

Towards a Discourse Theory of Abstracts and Abstracting

by

Timothy Robin Gibson, M.A., M.Phil.

**Thesis submitted
for the degree of
Doctor of Philosophy
University of Nottingham
May 1992**



Abstracts

Readers are invited to judge which abstract they prefer

Abstract 1

This Ph.D. thesis investigates the extent to which certain linguistic variables affect the perceived success of an abstract. In his opening chapter, the author situates the work as a contribution to discourse theory and formulates a set of seven basic research questions the study seeks to answer. Chapter 2 considers the most satisfactory design for the research. Part Two provides a survey of the relevant source material, consisting of reviews of: the Linguistics and Psychology literature; Artificial Intelligence computational summarising systems; and standards and guidelines taken from Information and Library Science. Part Three discusses the data collection: first, the collection of naturally occurring abstracts from a teaching establishment; second, the collection of judgements elicited by means of a set of questionnaires. The remainder of the thesis constitutes an attempt to reconcile these two types of data, words and opinions. The author draws a distinction between the qualitative and quantitative opinions of the judges (which he refers to as 'external measures') and the linguistic features present in the different abstracts ('internal measures') which determine these subjective opinions. Part Four discusses the data analysis, which draws on grammatical techniques from Systemic-Functional theory. In Part Four, hypotheses are investigated which relate the success of the different abstracts as perceived by the judges to the linguistic features present in the texts. Five different types of analysis are piloted using a small number of texts; three of these analyses are taken further and applied to all the abstracts. Part Five consists of two chapters. The first details the conclusions to be drawn from the study and explicitly answers the seven basic research questions introduced in Chapter 1. The second provides some suggestions for further research, chiefly concerning the collection of further external and internal measures. Finally, techniques from multivariate statistics are briefly sketched as a means of reconciling the two types of measure in the future.

Abstract 2

This Ph.D. thesis investigates the extent to which certain linguistic variables affect the perceived success of an abstract. More specifically, answers to seven basic research questions are sought. These include: what reasons do readers give for preferring one abstract over another?; is 'success' better explained by correlation with one, or with many, linguistic variables?; to what extent do readers agree with each other in their various preferences? and which linguistic features can help to explain readers' preferences?. In order to answer these questions, a total of 42 naturally occurring abstracts were collected from 29 second year Library and Information Science students at Brighton Polytechnic. 26 of these abstracts summarised General Knowledge source texts. 17 summarised Information Science (I.S.) source texts of three different types: journal articles, newspaper items and book chapters. The shortest I.S. abstract consisted of 111 words, the longest 651. Subjective data in the form of opinions of these abstracts were elicited by means of a set of six questionnaires. These questionnaires were administered to the students, to their two lecturers, and to 14 judges representing model consumers of such abstracts. The questionnaires elicited both quantitative and qualitative data. The 8 I.S. judges, for example, were asked to rank up to five different abstract versions summarising the same source text according to how helpful they believed them to be. They were also asked to provide reasons for their preferences. Different grammatical analyses from Systemic-Functional theory were employed to discover to what extent certain linguistic features in the abstracts determine their overall quality as perceived by the judges. Although some suggestions were made to overcome problems with the descriptive frameworks, analyses of generic structure and of cohesive harmony were found to be insufficiently reliable to enable precise hypothesis testing. However, the following linguistic phenomena were investigated more extensively and yielded interesting results: lexical texture; grammatical intricacy and choice of Theme. The answers to the above research questions are as follows. The reasons judges provide for preferring one abstract over another are many and varied; the two most common concern content and what might be termed 'reader-friendliness'. Success in text is a multivariate notion; any one linguistic measure cannot explain all the variation in judges' preferences.

Judges hold widely differing views of what constitutes a successful abstract: scores for Kendall's Coefficient of Concordance, w , a measure of inter-judge agreement, range from 0.109 to 0.597, suggesting that there are different drivers of success and that judges prioritise the importance of these drivers differently. In answer to the question, which linguistic features can help to explain readers' preferences?, the following results were obtained from the various hypotheses tested. Counter-intuitively, it was found that the more successful abstracts were characterised by lower levels of lexical density and were described as being 'clear'. Low levels of lexical density and lexical variation seem to be more the mark of 'reader-friendly' abstract writing, whereas higher levels of lexical density and lexical variation characterise abstracts which contain more information, but are correspondingly harder to process. In contrast to what is claimed in the literature, the hypothesis which stated that abstracts with a larger amount of clause level complexity would be generally preferred over abstracts with a smaller amount of clause level complexity was generally supported. Also, some clause combining strategies were noticeably preferred by the judges, while others were noticeably dispreferred. However, these preferences were not shared across the different abstract sets. Judges were found to be particularly sensitive to choice of Theme. A new type of Theme was identified to complement the two already existing sub-types of topical Theme, interactional and informational: Themes are to be regarded as *discoursal* if they refer to aspects of the source material, or to studies which are themselves discussed by the source material. Eight hypotheses concerning choice of Theme were investigated. For example, hypothesis 8 claimed that abstracts with more informational Themes would be preferred over abstracts with fewer informational Themes. This was supported for the Tanzanian set (H8c), but falsified for the other three. The judges seem to be indicating that they deem an informational style to be more appropriate for the Tanzanian source text. The three different types of topical theme serve different functions: informational themes primarily reflect the writer's desire to *enlighten*, by presenting the raw facts of the message for readers' consideration; discoursal themes primarily reflect the writer's desire to *orient* their readers, by providing a way of navigating through the various channels in which the information is presented; interactional themes primarily reflect the writer's desire to make it easy for readers to *integrate* the knowledge, by showing readers how the information relates to the various people involved in its transfer.

Acknowledgements

First and foremost, I would like to thank my supervisor, Margaret Berry, for all the help she has given me since I started as an undergraduate in the University of Nottingham. Margaret is the best supervisor one could wish for; I consider myself very lucky indeed to have had so many opportunities to benefit from her sound advice, clear teaching and unwavering support.

Ron Carter, Hilary Hillier, Gerald Parsons and many other friends in the Department of English Studies have also been a constant source of encouragement. I owe a special debt of gratitude to Caroline Stainton, who has been an enormous help throughout the different stages of the work.

One of the nicest things about trying to work out why some texts are perceived as being more successful than others is that inevitably one gets to meet lots of people. I should like to thank the students (some of whom may even now be writing abstracts professionally) and the other respondents to my various questionnaires. I think all these people would be surprised if they knew how much time I have spent deliberating over their work. I hope this may in part compensate them for the fact that I cannot thank them all by name.

Without Chris Turner, the Head of Department of Library and Information Studies, there would have been no field study at Brighton Polytechnic. I would also like to acknowledge the good-natured help and enthusiasm of Sharon Hall and Alex Simpson, the two lecturers referred to in Part Three below.

I am happy to acknowledge the encouragement of Chris Butler, Mark Caldicott, Andrew Hockley, Michael Hoey, Carol Maxwell, Nigel Pirie, Annabel Pollock, Rachel Ruffle, John Seton, Peter Thomas, Andrew and Stephen Thompson and Mark Velody, who provided support of a different kind which also helped to keep me going throughout the research. I owe special thanks to Sandra Williams and Dawn Wright, who worked with me on the automatic text summarisation project. I would also like to thank my parents for their patience and support.

I carried out this work while employed by BT firstly as an Artificial Intelligence programmer, and, more latterly, as a Human Factors engineer investigating quality of service, the measurement of customer satisfaction and other marketing concepts. Although at first sight these responsibilities do not seem to have very much to do with textlinguistics, I think they have all shaped the way I have carried out the research reported in this thesis. I should therefore like to thank BT not just for supporting this Ph.D., but also for providing me with such interesting and challenging work.

Table of Contents

Abstracts.....i
Acknowledgementsiv

Part One: Introduction

Chapter 1: Research Questions1
 1.1 Why ' ... abstracts and abstracting ' ?.....1
 1.2 Why ' ... a discourse theory ... ' ?2
 1.3 Why ' Towards ... ' ?.....4
 1.4 Orientation of this Thesis5
 1.5 Research Questions8
 1.6 Organisation of this Thesis.....9

Chapter 2: General Research Design 11
 2.1 Introduction 11
 2.2 The Multimethod Approach..... 13
 2.3 Data Collection..... 18
 2.3.1 Fieldwork 18
 2.3.2 Surveys 20
 2.3.3 Nonreactive Methods 22
 2.4 Data Analysis 23
 2.5 Conclusions..... 24

Part Two: Reviews of the Literature

Introduction	27
Chapter 3: Review of the Psychology and Linguistics Literature.....	29
3.1 Introduction	29
3.2 Some Brief Remarks Concerning Cognitive Psychology.....	29
3.3 Review of Work Relevant to Summarisation in Experimental Psychology	30
3.3.1 Sherrard 1986.....	31
3.3.2 Johns 1985	33
3.4 Review of Work Relevant to Abstracting in Linguistics	36
3.4.1 Nwogu and Bloor 1989	37
3.4.2 Drury 1989	39
3.5 Conclusions.....	42
Chapter 4: Review of Automatic Summarising Systems.....	44
4.1 Introduction	44
4.2 Introduction to the Shallower 'Spotting' Systems.....	44
4.2.1 An Example: Hoey's Lexical Connectivity Nets	45
4.3 Introduction to the Deeper 'Understanding' Systems.....	48
4.3.1 An Example: FRUMP	49
4.4 Conclusions	52

Chapter 5: Review of the Library and Information Science Literature..... 53

5.1 Introduction 53

5.2 Why Write Abstracts?..... 53

5.3 What are Abstracts?..... 54

5.3.1 What Sorts of Abstract are there?..... 55

5.3.1.1 The Needs of Readers..... 56

5.3.1.2 Type of the Original 57

5.4 How does One Write an Abstract? 57

5.4.1 Rules 58

5.4.1.1 The Topical Sentence 58

5.4.1.2 Structural Elements 58

5.4.1.3 Miscellaneous Recommendations 59

5.4.2 Procedures 60

5.5 Conclusions..... 61

Part Three: Data Collection

Introduction 62

Chapter 6: Fieldwork Data..... 64

6.1 Introduction 64

6.2 Objectives of the Fieldwork..... 64

6.3 Planning the Fieldwork..... 65

6.3.1 Planning the Ideal Site for the Fieldwork 65

6.3.2 Planning an Appropriate Role for the Fieldworker 67

6.3.3 Planning the Duration of the Fieldwork 69

6.3.4 Planning the Focus of the Fieldwork..... 69

6.4 The Brighton Polytechnic Case Study..... 70

6.4.1 The Abstract Writers 71

6.4.2 The Source Texts 72

6.5 Conclusions..... 73

Chapter 7: Survey Data..... 77

7.1 Introduction 77

7.2 Survey Data Requirements 78

7.2.1 The Requirement to Collect Source Material Process Data..... 80

7.2.2 The Requirement to Collect Source Material Product Data..... 81

7.2.3 The Requirement to Collect Abstracts Process Data 81

7.2.4 The Requirement to Collect Abstracts Product Data..... 82

7.3 Collecting the Data: Questionnaire Design and Administration..... 83

7.3.1 Selecting the Most Appropriate Informants to Question 84

7.3.2 Selecting the Most Appropriate Questions to Ask..... 85

7.3.3 Selecting the Most Appropriate Layout

and Means of Administration 94

7.3.4 Pilot Survey Data Collection 95

7.4 Conclusions..... 96

Chapter 8: Results of the Data Collection 98

8.1 Introduction 98

8.2 Source Process Data..... 99

8.3 Source Product Data101

8.4 Abstracts Process Data.....103

8.4.1 Students' Intuitions.....104

8.4.2 The Lecturer's Intuitions106

8.5 Abstracts Product Data111

8.5.1 Instructors' Opinions111

8.5.2 Producers' Opinions114

8.5.3 Consumers' Opinions115

8.5.3.1 Consumers' Ranks.....115

8.5.3.2 Consumers' Qualitative Remarks117

8.6 Conclusions.....121

Part Four: Analysis

Introduction	124
Chapter 9: Analysis of Generic Structure	131
9.1 Introduction	131
9.2 Why Analyse Generic Structure?.....	132
9.3 Problems with 'Genre'	133
9.3.1 A Summary Overview of Genre and Genre Study	133
9.3.2 Abstracts in Particular	136
9.4 Selection of Texts for Pilot Analysis.....	137
9.5 One Possible Interpretation of Distinct Personality Type, Abstract Version D.....	138
9.6 One Possible Interpretation of Distinct Personality Type, Abstract Version C.....	147
9.7 Discussion.....	152
9.8 Conclusions	153
Chapter 10: Analysis of Lexical Texture	154
10.1 Introduction	154
10.2 Lexical Texture as a Determinant of Success.....	154
10.2.1 Lexical Density	156
10.2.2 Lexical Variation.....	157
10.3 Statement of Hypotheses	157
10.4 Testing the Hypotheses	158
10.4.1 Reducing the Judges' Rankings to more Manageable Size	158
10.4.2 Correlating Observed and Expected Ranks.....	160
10.4.3 Deriving Quantitative Lexical Measures.....	161
10.5 Results of Testing the Hypotheses	165
10.5.1 Hypothesis 1 Results.....	165
10.5.2 Hypothesis 2 Results.....	170
10.6 Discussion of the Results	175
10.7 Conclusions	179

Chapter 11: Analysis of Cohesion181

11.1 Introduction.....181

11.2 Cohesive Harmony.....183

11.2.1 Stages in the Analysis183

11.2.1.1 Lexical Rendering184

11.2.1.2 The Formation of Chains.....186

11.2.1.3 Plotting Chain Interaction.....193

11.3 Static and Dynamic Ties197

11.4 Conclusions199

Chapter 12: Analysis of Grammatical Intricacy.....201

12.1 Introduction.....201

12.2 Readability and Syntactic Maturity as Determinants of Success.....202

12.2.1 Simple Readability Formulae203

12.2.2 Formulae Based on Surface Structures.....204

12.2.3 Formulae Based on Derivational Complexity.....208

12.3 Halliday's Conception of Grammatical Intricacy.....209

12.4 Clause Combining as a Determinant of Grammatical Intricacy210

12.4.1 Clause Complex Analysis211

12.4.2 Criticisms of Clause Complex Analysis.....216

12.5 Statement of Initial Hypothesis.....218

12.5.1 Testing the Initial Hypothesis.....218

12.5.2 Problems of Analysis.....219

12.5.2.1 'Ungrammaticality'.....220

12.5.2.2 Segmentation222

12.6 Results of Testing the Initial Hypothesis.....224

12.7 Further Hypotheses.....229

12.7.1 Testing the Further Hypotheses.....229

12.7.2 Problems of Analysis.....230

12.7.2.1 Determining Interdependencies.....230

12.7.2.2 Determining Logico-Semantic Relations.....234

12.7.2.3 Measuring how Clauses Combine235

12.8 Results of Further Testing238

12.8.1 Tanzania Results.....240

12.8.2 Distinct Personality Type Results.....244

12.8.3 Vickery and Vickery, Chapter 1 Results246

12.8.4 Vickery and Vickery, Chapter 2 Results248

12.9 Further Discussion of the Results250

12.10 Conclusions.....252

Chapter 13: Analysis of Themes	254
13.1 Introduction.....	254
13.2 Why Analyse Themes?.....	255
13.3 Theoretical Background.....	256
13.3.1 Halliday's Approach	256
13.3.2 Developments of Halliday's Approach	258
13.4 Statement of Initial Hypotheses	259
13.5 Testing the Initial Hypotheses	261
13.5.1 Problems of Analysis Arising from the Nature of the Data.....	262
13.5.2 Problems of Analysis Arising from the Imperfections of the Descriptive Framework.....	264
13.5.2.1 The Delimitation of Theme.....	264
13.5.2.2 Theme in Non-Main Clauses	267
13.5.2.3 Ellipsis	272
13.5.2.4 Grammatical Metaphor	273
13.5.2.5 Some Remarks on Marked Theme	276
13.5.3 Summary of the Decisions Taken in the Light of these Problems	278
13.5.4 Deriving Quantitative Measures to Test the Initial Hypotheses	280
13.5.5 Results of Testing the Initial Hypotheses	287
13.5.5.1 Hypothesis 1 Results	289
13.5.5.2 Hypothesis 2 Results	300
13.5.5.3 Hypothesis 3 Results	307
13.5.5.4 Hypothesis 4 Results	313
13.5.5.5 Hypothesis 5 Results	316
13.5.5.6 Conclusions from the Initial Hypotheses	321
13.6 Further Hypotheses.....	323
13.7 Testing the Further Hypotheses.....	325
13.8 Results of Testing the Further Hypotheses	335
13.8.1 Hypothesis 6 Results.....	336
13.8.2 Hypothesis 7 Results.....	336
13.8.3 Hypothesis 8 Results.....	337
13.9 Discussion of the Results	339
13.10 Conclusions.....	342

Part Five: Conclusions and Suggestions

Introduction.....	346
Chapter 14: Conclusions.....	347
Research Question 1.....	347
Research Question 2.....	349
Research Question 3.....	351
Research Question 4.....	352
Research Question 5.....	353
Research Question 6.....	354
Research Question 7.....	356
Lexical Density and Lexical Variation.....	356
Grammatical Intricacy.....	357
Theme Choices.....	359
Chapter 15: Suggestions for Further Research.....	364
15.1 Introduction.....	364
15.2 Collecting Further External Measures.....	364
15.2.1 Collecting Further External Qualitative Data: the Grocery List.....	365
15.2.2 Collecting Further External Quantitative Data.....	367
15.3 Collecting Further Internal Measures.....	369
15.4 Reconciling the Two Types of Measure.....	374
15.4.1 Factor Analysis.....	375
15.4.2 Discriminant Function Analysis.....	377
References.....	382

1

Research Questions

1.1 Why ' ... abstracts and abstracting ' ?

The aim of this study is to investigate the extent to which certain linguistic variables affect the perceived success of an abstract. The notion of success will be returned to in the next section. The abstract as a particular type of text will be familiar to most academic readers - many will have prefaced their own journal articles with abstracts, or will have sent them to conference moderators - but the abstracts investigated in this thesis are not written by the authors of the source material for which the abstracts are designed to be surrogates, but are composed by third-party writers. Both author-prepared and third-party abstracts are to be found in abstracting journals; they can also be retrieved from online textual databases. As technology progresses, researchers gain access to more and more information, much of it in text form. But having too much information is often as unsatisfactory as having too little, which is why the abstract has developed as a means of saving readers' time by helping them prioritise material to be read.

The work described in this thesis was carried out with two separate, though related, aims in mind. The motivation underlying the first of these came from business and industry, more specifically from BT's research initiative in Natural Language Processing, one part of the company's commitment to exploit the commercial promise of Artificial Intelligence. The project brief was to write computer software which would automatically generate abstracts from written inputs of arbitrary length, subject matter and complexity. At the beginning of the project many existing automatic abstracting machines were evaluated and it quickly became apparent that even the most carefully-engineered systems produced output that was incoherent or otherwise unsuccessful; further, there was nothing in the literature to suggest how an abstract had to be constructed so that it would be perceived to be a successful piece of text in its own right. This thesis does not report on the summarising

software which was written during the project, but rather documents the research that was undertaken to provide a principled foundation for the design of a successful summarising system. The motivation underlying the second aim of the research was educational, and stemmed from the desire to help human writers compose abstracts which would satisfy their readers. This thesis does not, however, constitute a prescriptive 'How to write good abstracts' manual. Although many such manuals have been published, often these contain little of any practical linguistic substance, and, where guidelines are made sufficiently explicit to be acted upon, there is no guarantee that following the rules will result in writing judged to be of sound quality by readers. By investigating how and why some abstracts are preferred over others on real occasions on the other hand, it is hoped that this thesis will go some way towards laying a firmer foundation for the design of more linguistically-motivated teaching materials which may engender better quality abstract writing.

But why abstracts *and* abstracting? There is a tendency, especially when one's object of enquiry is writing, to conceive of textual data only as a product, divorced from its means of production and reception. The text in its guise as artefact is of course very worthy data and has been studied for many centuries, but there are growing compensatory moves in textlinguistics to concentrate more on the underlying system, in order to show how issues of process determine the character of the product. Although this thesis does employ descriptive apparatus founded on static rather than dynamic grammatical models, there are also attempts to elicit informants' intuitions concerning process-type phenomena.

Informants of different sorts play a very large part in giving direction to this thesis. The next two sections discuss the idea of success, and why it was considered necessary to enlist informants' judgements.

1.2 Why ' ... a discourse theory ... ' ?

The phrase 'discourse theory' comes from Morgan and Sellner (1980: 196) who say 'The proper business of a discourse theory is to say what makes a *text* successful or not, clear or not, coherent or not, in terms of the relation between the text and its intended content' (original emphasis). This point is well taken in this thesis.

Judging the quality of language is not altogether unlike the process by

which business and industry judge the quality of the products or services they offer their customers. Manufacturers of toasters, for example, know that two types of measures can be taken: the first kind are 'internal', and pertain to the workings and efficiency of the toaster - how much it costs to produce, what its life expectancy is, how long it takes to cook the toast, and so on; the second kind are 'external', and pertain to the consumer's perceptions of it - whether purchasers like it, whether they think they are getting good value for money, what the likelihood is of them buying more products from the same manufacturer, and so on. It is clear that these two kinds of measure are very different.

In effect, Morgan and Sellner's conception of discourse theory is more an internal definition than an external one; by concentrating on the 'what', they have relegated the matter of the 'who': whose intuitions of success, clarity, and coherence are to be accounted for? The approach taken in this thesis is that the best people qualified to have such intuitions about abstracts are those people who regularly make use of abstracts in their everyday lives. Once the 'external' measures have been properly collected, it is the linguist's job to attempt to explain these intuitions using whichever 'internal' measures are deemed appropriate. It is only by attempting to reconcile these two types of measure that real progress will be made towards explaining why some texts are perceived to be more successful than others. Enkvist (1978: 6) describes very well what he believes to be the responsibility of the linguist:

'As ordinary users of language, we are all familiar with phrases such as "smooth flow of discourse" or "tight coherence" or "loose and rambling text". That such phrases are used suggests that they fill a need. We should therefore not wrinkle our noses at them and dismiss them as impressions which defy description in properly linguistic terms. On the contrary, here - as always - a linguist's job should be to correlate intuitions with specific linguistic forms and functions. By so doing, the linguist may try to divest such impressionistic terms of needless mystique and see to what extent smoothness, tightness, looseness and other such textual attributes can be found to correlate with precisely describable linguistic phenomena'.

This is precisely the view taken in this study: ordinary users' qualitative opinions, although quite possibly vague, are sought out and valued. It is hoped that divestiture of impressionistic terms can then be achieved by matching together the two types of measure: external measures, which seek to describe informants' attitudes, enable texts to be compared and contrasted in terms of their differing levels of perceived quality; internal measures, which seek to describe textual phenomena, enable explanation of the external measures, by identifying features of the discourse which give rise to judges' preferences.

1.3 Why 'Towards ...' ?

There are two reasons why the word 'Towards' is included in the title of this work. One is to acknowledge the debt to Sinclair and Coulthard, whose book published in 1975 was called 'Towards an Analysis of Discourse'. Sinclair was one of the first workers in the field to state that one of the aims of linguistics should be to specify the circumstances under which particular utterances would be judged successful (reported in Dixon 1967: 76, cited by Berry 1989). Sinclair and Coulthard 1975 has had a profound influence on discourse analysis, and their concerns with success are shared by many Systemic linguists; Halliday, for example, has written (1985: xv) that 'The higher level of achievement is a contribution to the evaluation of the text: the linguistic analysis may enable one to say why the text is, or is not, an effective text for its own purposes - in what respects it succeeds and in what respects it fails, or is less successful'.

The second reason is that, like Sinclair and Coulthard, this work represents only a first step along the road towards discovering why some texts are considered more successful than others. Sinclair and Coulthard's major contribution was to provide a practical descriptive framework for analysing spoken discourse; this thesis aims to provide a practical descriptive framework for analysing success in abstract writing by identifying, firstly, which features of the language genuine consumers of abstracts think differentiate good examples from bad, secondly, to work out which of these many different factors seem to matter most, and thirdly, to provide some linguistic underpinning to those opinions which are expressed only impressionistically.

The word 'Towards' also brings to mind Stubbs' observations that discourse analysis is, firstly, a new area of study (1983: 129), and secondly, at a very elementary stage (1983: 86). Partly because of this, Stubbs exhorts his readers to think carefully about what kinds of data can and should be explored, to think carefully about the relationship between data and theory, and, more generally, to combine tools and techniques. The next chapter, Chapter 2, is principally about methodology, but it was thought appropriate that this first chapter should provide something of an overview of the central tenets of the thesis in which Stubbs' suggestions are considered and other 'operating principles' are made clear. This is done in the next section.

1.4 Orientation of this Thesis

In an interesting and informative paper (1976) comparing and contrasting the research styles inherent in discourse analysis (DA) and in textlinguistics (TL), Sandulescu defines a number of key differentiating concepts, discussing each paradigm's attitude to each. Although not all of the seventeen propositions he identifies are relevant to this study, the paper will be used to derive a set of underlying principles which should help define the orientation and general working assumptions of this thesis.

Firstly, Sandulescu makes the terminological point that what is called 'discourse' by English and American scholars is more usually called 'text' by scholars on the Continent. Since both schools have much to recommend them, this thesis will use the terms interchangeably.

Sandulescu's second proposition (1976: 351) states that whereas DA is fundamentally a data-centred endeavour, TL is more a model-oriented approach. The present study, like DA, is *data-driven*, rather than *theory-driven*. There are many reasons for this, but chief amongst these is the fact that there does not yet exist, as Part Two below aims to demonstrate, a satisfactory or even promising theory which can explain why some abstracts should be considered more effective than others. Therefore it would seem sensible to seek answers by starting with the data itself.

Sandulescu's third and fourth propositions (1976: 351) concern the kinds of result each school expects to attain. DA expects to further knowledge by collecting (classified) instances of language use, whereas TL expects to further knowledge by making testable generalisations. Partly because so little is known about abstracting practice, this study is more *naturalist* than *positivist* (Hammersley and Atkinson 1989: 3 - 9). Because much importance is attached to collecting realistic, substantive data, this study aims to accurately identify and investigate, without too much disturbance, certain particularised instances of language use, rather than to posit explanatory laws at the outset which are necessarily generalisable to other situations. Generalisation cannot realistically be attained in an exploratory work such as this, but because generalising to other situations is an ultimate goal, hypotheses will be formulated in a way that can facilitate testing in different contexts.

Sandulescu's fifth proposition (1976: 352 - 353) states that whereas DA subscribes to the view that structure at the discourse level cannot be accounted for by the mechanisms which have developed for describing and explaining

structure at the grammatical level, TL subscribes to the view that texts can be described and explained by suitably enriched sentence grammars. The view taken in this thesis represents something of a mixture of these two positions: here it is argued that certain types of grammatical apparatus can be used to demonstrate why written texts are more than just a random collection of written sentences, depending of course upon one's conception of the role and scope of 'grammar'.

Unlike TL, which places great store by the deep structure versus surface structure distinction, and by the competence versus performance distinction, both guiding principles of Chomskyan thinking (see Sandulescu's propositions 6 - 11, 1976: 353 - 356), the present study is more *functional* than *formal*. This is because it was thought that a more functional approach to abstracting would better show how abstracts are used, how some fulfil informational and aesthetic requirements, and how others do not.

Not surprisingly, much of Sandulescu's paper centres around discussion of the type and status of the data in the two schools (propositions 7 - 14, 1976: 353 - 358). Because the present work must be regarded as more *exploratory* than *confirmatory*, it was thought more appropriate to adopt a more *qualitative* than *quantitative* approach. This should not be taken to mean that the study concerns words not numbers, but rather reflects an attitude to their relative importance and interrelationship. In this thesis calculations are carried out only to the extent that they may more easily falsify hypotheses relating judges' observed preferences to those predicted by particular linguistic features of the data. The qualitative style of research also implies a particular attitude to the type of data used. Much of the data on which the work is based has been studied *in vivo*, rather than *in vitro* (Givón 1979: 22 - 26), that is to say in the field rather than in the laboratory, or in the armchair. This helps to ensure that the textual data collected will be a reasonable representation of those naturally occurring abstracts written by real abstractors on real occasions for real readers with real needs. Similarly, use of the analyst's introspections has been largely resisted so that the attitudinal data collected will be a reasonable representation of the opinions of genuine consumers of such abstracts. This should not be taken to mean that no use is made of intuition - the intuitions of the consumers of abstracts are sought out and valued - it simply means that the researcher/analyst's *own* success judgements were not considered a sufficient source of data for the study.

Although Sandulescu does not make this particular point, it could be argued that DA is more *inductive* than *deductive*, whereas TL is more

deductive than inductive. Hypothetico-deductionism is more appropriate for transformationalists because they have chosen a priori to focus upon competence data, deliberately ruling out those variables they believe to be extraneous (such as ontogenetic factors, error phenomena, the function of the language, and so on). However, since this is pioneering work in that it was not known at the beginning of the study which variables were most relevant to success, it was thought preferable to proceed inductively rather than deductively. The main advantage of inductive thinking is that it generally allows more freedom, and encourages a more flexible attitude to change as patterns in the data become evident. This seems to be Stubbs' preferred strategy: 'What is needed is a balance in which *emergent* theory is controlled by data' (1983: 129, emphasis added). The main advantage of deductive thinking on the other hand is that hypotheses are clearly stated and are rigorously subjected to falsification. This thesis aims to combine the strengths of both types of discovery procedure.

Sandulescu is clearly very interested in how language is used, and believes practitioners of both DA and TL should be too, a bias which is evident in many of his propositions. The present study focuses very sharply on language use, and, as has been discussed above, has *applied* rather than *theoretical* objectives. This is not to say there is no theoretical discussion in this thesis - there is much - but theoretical issues are only deliberated when they affect the use or interpretation of particular linguistic means of describing regularities in the data.

In a nicely ambiguous concluding section Sandulescu observes (1976: 362) that one paradigm 'caters for highbrow linguistic tastes, whereas the other approach is quite aware of the fact that it aims primarily at the lowbrow'. If work which is applied, data-centred and functionally oriented can indeed be described as 'lowbrow', then this thesis must also be evaluated in such a light. Again in his summing up, Sandulescu observes that DA is eclectic, whereas TL is elitist in its tolerance of other approaches. Although the model adhered to in this thesis is firmly systemic-functional, other types of linguistic description have been explored, and methods from disciplines other than linguistics have been imported. This was done to try to achieve a more rounded investigation. This and many of the above points will be returned to in the following chapter.

These various assumptions have been deliberately worded using 'rather than ...' and 'more *x* than *y*' formulations to signal relative preference, rather than absolute choice. For example, the fact that the study is more inductive

than deductive should not be taken to mean that the thesis contains no deductive-type argument; as mentioned above, explicit hypothesis testing is carried out in the analysis phase of the study. In fact, much if not all of the work presented in this thesis has been conducted in a multimethod style, which, if followed to the letter, obliges researchers to ponder the advantages and disadvantages of each and every method they propose to adopt, and then to attempt to combine those with complementary strengths and non-overlapping weaknesses. This procedure and its appropriation into what follows will be returned to in the next chapter.

Neither should these tenets be construed as what Wolcott (1990) calls 'academic throat-clearing'. They are not meant to declare allegiance to a particular school, nor should they be taken to constitute any kind of a disclaimer. Rather they are stated at the outset to act as orientation for the reader, to show what kind of study follows, and to foreshadow much of the spirit in which later discussion is presented.

The next section, section 1.5, discusses some of the questions which this thesis was intended to answer, and section 1.6 below concludes this introductory chapter by discussing the organisation of the rest of the thesis.

1.5 Research Questions

The process of research invariably starts with a set of questions or issues, or, to use Malinowski's preferred phrase (1922: 9), 'foreshadowed problems'.

Formulating such questions at the outset has two important functions, one of which is to the advantage of the researcher, the other to the person reading the results of the research: firstly, the basic research question 'identifies the phenomenon to be studied. It tells you [the researcher] what you specifically want to focus on and what you want to know about this subject' (Strauss and Corbin 1990: 38) - this should be used to design the study (see next chapter), and can be returned to throughout the process of the research for inspiration, clarification or revision, if necessary; and secondly, a clear statement of the basic question or questions can help orient readers to the aims and objectives of the research by situating the work as a response to a known problem situation.

In spite of these benefits, researchers often find that committing these questions to paper is a very hard thing to do. In this case, for example, it has not been possible to crystalise the research in one single question, although the very first sentence of this thesis - 'The aim of this study is to investigate the

extent to which certain linguistic variables affect the perceived success of an abstract' - captures one of the basic goals, if not the spirit, of the investigation.

The following, therefore, bearing in mind Miles and Huberman's admonition (1984: 36) that 'formulating more than a couple of dozen general questions is looking for trouble', is not one single question but a shortened list of some of the most important questions this research was designed to address:

- What reasons do readers give for preferring one abstract over another?
- Are all of these readily interpretable? In other words, can they all be used to suggest improvements in a straightforward way, or are some less 'actionable', or, to use Enkvist's term quoted above, are some more 'impressionistic'?
- If readers do give impressionistic reasons, how should these be related to, once again in Enkvist's words, 'precisely describable linguistic phenomena'?
- Is 'success' better explained by correlation with one, or with many, linguistic variables?
- If success is determined by more than one linguistic variable, is it possible to say which variables are most/least important?
- Do readers think alike? More specifically, to what extent do they agree with each other in their various preferences?
- Which linguistic features can help to explain readers' preferences?

There is a last, over-arching question, which is: how best can the above questions be investigated?. In answer to this, research design and methodological issues are made the subject of the next chapter.

1.6 Organisation of this Thesis

This thesis contains five major parts, each of which, except the first, has a short introduction. Part One continues after this chapter with a discussion of

how the research was designed.

Part Two is devoted to surveying the relevant literature, and contains reviews of work in Linguistics and Psychology (Chapter 3), work in Artificial Intelligence (Chapter 4), and of work in Information and Library Science (Chapter 5).

Part Three covers data collection: Chapter 6 describes how naturally occurring data was collected in the field; Chapter 7 describes how this was supplemented by data elicited by survey instruments; and Chapter 8 brings these together by detailing the results of the data collection.

Part Four covers the analysis phase of the study and is meant to provide a reconciliation between the two types of data, texts and success judgements, collected in Part Three. A number of different linguistic techniques are used to demonstrate the degree of correlation between the informants' judgements and the specific linguistic forms and functions of the abstracts: Chapter 9 examines generic structure; Chapter 10 investigates the abstracts at the lexical level by measuring their lexical density and variation; Chapter 11 moves onto their cohesive properties; Chapter 12 assesses their grammatical complexity by considering their clause complex structure and embedding; and Chapter 13 is devoted to choice of Theme. The reasons why these particular types of analysis were chosen are fully documented in the Introduction to Part Four, but, very briefly, the choice of these linguistic tools was motivated by the judges' more impressionistic remarks. For example, much of the qualitative data collected makes reference to issues of coherence. The testing reported in Chapters 10 and 12 was therefore carried out because one suggested way of explaining coherence is in terms of 'texture', which Halliday believes (1985: 313) can be created by the careful arrangement of thematic structure and cohesion.

Finally, Part Five provides conclusions and suggestions for further research.

2 General Research Design

2.1 Introduction

This chapter considers methodological issues in so far as they relate to the basic aims and objectives of the research, as set out in Chapter 1. The following discussion introduces the general methodological orientation of the study, and shows how the research programme was designed in order to shed the most light on the research questions stated in the first chapter. The comments made in this chapter are of a general nature; Chapters 6 and 7 below, which report on the data collection stages, detail the rationale underlying specific decisions taken once the basic method was decided upon.

Writing about common practices in sociological research, Silverman (1985: 13) observes that readers are oriented to the basic premises of the work typically in one of two ways. Either the researcher can declare allegiance to a particular school, and thereby take its associated goals, beliefs, methods, and so on, as read, or s/he can issue a disclaimer, such that the study is presented as being essentially descriptive, or exploratory, and therefore not subject to theoretical concerns. In both cases methodological issues are skirted, and Silverman's point is that neither extreme is entirely satisfactory.

More preferable would be to adopt facets of each approach (the idea of having the best of a number of different worlds will be discussed further in section 2.2 below) such that researchers *ground* their work with respect to the larger theoretical situation, while still concentrating their interest on their own particular subject of enquiry.

In his book *Schools of Linguistics* (1980) Sampson seems to bemoan the fact that many linguists feel obliged to define their own thinking in terms of Chomsky's position (page 130 f.f.). In fact, relating one's work to other people's is often a useful procedure, as it can help establish what one is trying to do, as well as what one is not trying to do, and how one proposes to go about it. In a paper entirely devoted to methodology written for Systemic linguists, Berry

1989b uses this technique to draw out the similarities and differences concerning the status of *data* between, on the one hand, the Chomskyan camp, and Systemicists on the other. In so doing, she is able to demonstrate how Systemicists could usefully take over some of the methodological principles of the transformationalists, while still retaining fundamentally different goals.

The data that transformationalists use is of two types: sentences and judgements, such that each sentence is paired with a judgement as to its grammaticality (Labov 1972b: 186). The theory that transformationalists formulate attempts to reconcile these two types of data; to discover why some sentences are judged grammatical, and some ungrammatical. Berry's point (1989b: 11 f.f.) is that this way of investigating syntactic phenomena has relevance for the higher level of discourse: whereas transformationalists concern themselves with sentences, researchers interested in discourse concern themselves with texts (a difference of scale, 1989b: 13); and whereas transformationalists make up their own sentences to be the object of enquiry, researchers interested in discourse ultimately want to be able to make claims about how language is used in the real world, and so prefer to investigate naturally occurring data whenever practical (a difference of naturalness, 1989b: 16).

The research goal of the present study is to discover why some abstracts are judged more successful than others, and so it is to be expected that the data (in the form of the words) will differ from that of transformational theory in both scale and naturalness. However, it is also to be expected that the data (in the form of the judgements) will differ from that of transformational theory in other ways too. Firstly, whereas transformationalists are principally interested in grammaticality, the fundamental research aim here is to account for 'well-formedness' at a much higher level. Secondly, whereas transformationalists assume that every native speaker of a language can act as a legitimate informant for well-formedness at the syntactic level, no such assumption can be made in this study. The transformational notion of the 'ideal speaker-listener, in a completely homogenous speech-community' (Chomsky 1965: 3) effectively places few restrictions on the type of person who can act as a judge, but here the situation is different. It is likely that, unlike Chomsky's ideal speaker-listeners, different judges may well have different criteria for success, so stricter restrictions must be placed upon the selection of judges. Care must be exercised to ensure that opinions are sought from people who form a representative sample of the population to be accounted for, an argument the ethnomethodologists would sympathise with, see for example

the work of Sacks and Schegloff; in the case of the present study, judges would comprise those people who make use of abstracts as a part of their everyday working lives.

So far this discussion has concentrated on data. It has been argued that while there are differences between the approaches in terms of the type of data used, the most important of which being naturalness, both in the form of 'natural texts' and 'natural judges', there is a similarity in terms of the status accorded to data in both the paradigms (see also the comments made above in Chapter 1 concerning induction versus deduction). The interrelationship of theory and data will be taken up in the next section, and the idea of designing naturalistic research will be further explored.

2.2 The Multimethod Approach

This section discusses why a multimethod approach was judged to be the most appropriate type of research methodology for the present study, given the aims and objectives stated in Chapter 1 above. The multimethod approach is an eclectic style of research, and, as is argued below, provides a powerful design for investigating issues within discourse theory. In what follows, a word-processing analogy is used to introduce the fundamental concepts of the multimethod style; these are then further developed in discussions of data collection (section 2.3) and data analysis (section 2.4).

For most people writing a Ph.D. is an extremely hard task. Demands are placed on candidates at a number of different levels, each with its own type of problem. One obvious difficulty is the sheer physical volume of words to be written, only a small fraction of which will find their way into the final thesis.

This problem can be made easier through the careful use of a word processor, for example. However, the choice of word processor to be used will depend to a large extent on the problem the word processor is intended to solve. Different problems necessitate different solutions; for instance, it is likely that the person writing up a Ph.D. in Engineering will include many complex algebraic equations, and so will need a machine capable of writing symbols not normally encountered in more everyday fields. Clearly then, the first thing to do is to formulate a list of requirements or features that will help provide a solution to the problem. For a word processor these features might include: the ability to incorporate graphics with text; a spelling and thesaurus facility; a high degree of usability; speedy response; mail merge; automatic help with the

construction of indexes, contents pages, references, and so on.

However, it is extremely unlikely that any one word processor will realize all the desired features in an entirely satisfying way, so a choice must be made to pick the one which solves the most pressing problems. Typically users will prioritise their requirements, and then attempt to match these against the facilities offered by the competing systems, often taking into account other factors such as cost, availability and so on. It is inevitable that some compromises will be made at this point, since it is often the case that word processors that are particularly good in some areas are poor in others. In the end though, one should hope to end up with the word processor that is most appropriate for one's overall goals.

There is, however, an alternative strategy. One could elect to use a combination of different word processors, using the best features of each. Under this regime there is a better chance of achieving exactly what is required, since one is no longer subject to the limitations of any one system. This is the central tenet of the multimethod style of research, which is a 'strategy for overcoming each method's weaknesses and limitations by deliberately combining different types of methods within the same investigations' (Brewer and Hunter 1989: 11). It is argued that such diversification has a number of theoretical advantages, but it must be remembered that there are a number of resulting disadvantages. In the word processing scenario these concern practical problems of resourcing, chiefly in terms of time and money. First of all, one has to spend a great deal more money, buying perhaps six sets of software instead of one. Secondly, one has to spend a great deal more time, learning how to use each system. Thirdly, and perhaps most seriously, each has its individual quirks; further, there are often incompatibilities between different packages, and these make the production of a fully-integrated and satisfying piece of work problematic.

Moving on now from word-processing to linguistics, researchers interested in how language is used will already have met the multimethod concept in the guise of *triangulation* (see for example Cicourel 1964), and in the work of Labov more generally (see especially Labov 1972). But while Stubbs (1983: 235), for example, is careful to point out that triangulation should not only be thought of in terms of measurement, and should be extended to cover data collection, few linguists have taken seriously Brewer and Hunter's contention (1989: 124) that the multimethod approach 'is a perspective that permeates all stages of the research process from initial theoretical hunches to final publication'.

One of the motivations underlying the remaining sections of this chapter

then is to demonstrate why multimethod design was thought appropriate to address the basic research questions of the present study, and to point out the stages of the work in which it was exploited. Section 2.3 below will give an overview of the strengths and weaknesses of the different types of data collection which can be adopted within a multimethod investigation, and will provide pointers to the chapters which describe the specific details of how the data was collected in the study. Section 2.4 will pave the way for the later chapters in Part Four by showing how multimethod principles can be used in the analytic phases of an investigation.

Before discussions of data collection and data analysis take place, however, the multimethod perspective will be used to explain more deeply the relationship between the data and theory in this study.

The concepts of theory and data entail questions of scientific method in general, and the relationship between theory construction and theory testing in particular. Transformationalists test their theories by formulating hypotheses, and then attempting to falsify them through the use of counter-examples. This is essentially a hypothetico-deductive approach (Popper 1959) and is characteristic of what many people believe to be *the* scientific method. Berry 1989b argues that Systemicists can benefit from being more Popperian in their methods, and in this thesis every attempt will be made to provide clearly stated hypotheses and to subject these to rigorous testing.

Where the hypothetico-deductive method is less useful, however, is in the stage prior to hypothesis testing, namely hypothesis generation (Becker 1970: 21; Bulmer 1984b: 248). Thanks to the homogeneity principle referred to above, transformationalists feel they can make heavy use of introspection (whether this is indeed justifiable is discussed in Givón 1979), so much so that there is no clear separation between the two stages. But this discovery procedure is not applicable to the aims and objectives of the present study, since it has already been argued that the intuitions to be accounted for should not be assumed to be those of the analyst, but rather those of the relevant abstract writer-readers in the field.

Just as there are pros and cons associated with hypothetico-deductionism, so there are pros and cons associated with analytic induction, another type of scientific method. The advantages of analytic induction which are particularly relevant for this study are as follows. First of all, analytic induction stresses the need to remain 'open' to the data (Bulmer 1984b: 255). One of Givón's criticisms of the transformationalists is that they have allowed their theory to define (he would say 'curtail') *a priori* their data, such that all the variables

which are not believed to have a direct bearing on grammatical competence (such as the structure of the language processor, the effects of diachronic change, and so on) are deliberately excluded. Whether or not this limiting of the investigation is a legitimate enterprise for transformationalists is a moot point. In the case of abstracting, on the other hand, one of the objectives of this research is to find out what these variables might be, so it would not make sense to exclude any factors at the beginning of the investigation without knowing more about how exactly they might affect perceived levels of success. In this way, analytic induction is more driven by data, and less oriented toward any particular theoretical model. Secondly, the data that is used to drive the investigation is most usually natural, in that analytic induction promises to make sense of field data (Lincoln and Guba 1985: 202) by studying it *in vivo* (Glaser and Strauss 1967: 40), taking the perspective of the actual participants involved in the interaction. Lastly, one of the primary functions of analytic induction is to generate hypotheses (Marshall and Rossman 1989: 43 - 44); these are expected to emerge throughout the course of the inquiry, rather than to magically appear at the outset. In this last connection it is perhaps worth noting that methods themselves are sometimes expected to emerge throughout the course of the inquiry, and that many inductionists would argue that it is not in fact possible to specify a rigid design prior to field exposure. This is doubtless one of the reasons why it is not always possible to provide a methodology for discourse analysis in a 'recipe-style format' (Potter and Wetherell 1987: 158).

What are claimed to be strengths of induction are sometimes criticised as being weaknesses, however. The flexibility of emergent design can sometimes give rise to studies that are unstructured, or otherwise unconvincing, though Miles and Huberman (1984: 20 - 21) are quick to point out that analytic induction should not be taken to mean that 'anything goes', and that it is perfectly possible to be both inductive and structured. Another peculiarity of analytic induction is that the normal requirement that a problem situation is precisely stated at the beginning of the research (as was carried out in Chapter 1) is relaxed, such that this too can be allowed to evolve throughout the course of the work. The issue of testing is addressed by inductionists, although somewhat naively, since further observation is typically used to verify, rather than to falsify, the propositions put forward by the theory, a stance severely criticised by Popper (for example, 1976: 78 - 87). Further, it must be admitted that the 'propositions' are not always postulated in the form of readily testable hypotheses, another potential barrier to scientific advancement.

There are things to be admired then both in the deductive strategy, and in the inductive strategy. While there are explanations as to why such little progress has been made to achieve symbiosis between the two positions (Sjoberg and Nett 1968: 35), there seem to be no real theoretical reasons why certain facets of the two should not be combined; indeed many researchers argue that this is absolutely necessary (Bulmer 1984a: 4).

Lachenmeyer neatly motivates a multimethod perspective when he says 'Theories are not developed deductively or inductively but *both* deductively and inductively. There is a constant interplay between the observation of realities and the formation of concepts, between research and theorizing, between perception and explanation. The genesis of any theory is best described as a reciprocal development of observational sophistication and theoretical precision' (1971: 61, original emphasis). With regard to the research presented here, only naturally occurring abstracts have been collected; further the opinions, judgements and experiences of both genuine producers and consumers of abstracts have been sought in as naturalistic a way as possible (Chapters 6 and 7). These tasks have been conducted to ensure a high degree of observational sophistication. The problem situation has been clearly stated at the outset (Chapter 1), and the two types of data (the naturally occurring abstracts and the opinions) will be seen to be reconciled in such a way that testable hypotheses could be made, and could be subjected to falsification (Part Four). These tasks have been conducted to ensure a high degree of theoretical precision.

The remainder of this chapter will demonstrate that the reconciliation between naturalistic observation and theoretical testability can best be brought about by combining both qualitative and quantitative methods (Kirk and Miller 1986: 70). By the very nature of these two adjectives it is tempting to suppose that the fundamental difference between the techniques comes down to whether or not the analyst proposes to *count* certain features of the data. But this is overly simplistic; qualitative work is better conceived of as a commitment to observe, in a variety of different ways, people doing everyday things, in their own everyday worlds. This high degree of naturalness does not of course preclude counting, but the fact that qualitative researchers make deliberate attempts to *control* the situation as little as possible often means that the effects of the variables that they are interested in become correspondingly less clear.

In practice, qualitative research techniques are particularly appropriate when 'there is a need to explore interactions among ambiguous or unclear

variables' (Marshall and Rossman 1989: 42), or, as is the case in this particular study, when there is a need to discover what the variables themselves might be. Once known, analysis can then take place so as to investigate more fully what effects these variables have. As will be seen below, it is at this stage typically that a quantitative style of research becomes more appropriate.

2.3 Data Collection

Three different basic methods, or styles, of data collection have been adopted in this study, so as to 'partially overcome the deficiencies that flow from one investigator and/or one method' (Denzin 1970: 300). These three styles are fieldwork, surveys, and nonreactive methods; these will be discussed in the three subsections that follow. There is a fourth style of data collection, which is experimentation. The present study is meant to act as a precursor to a follow-up experimental stage, in which well-motivated hypotheses are investigated in more controlled conditions. Experimentation is beyond the scope of the present work, however; but possible types of experimentation are discussed in Chapter 15 below.

2.3.1 Fieldwork

Chapter 6 below provides a catalogue of all the fieldwork data collected for the study, and documents the planning that was carried out prior to the researcher setting out into the field. This section will briefly summarise why fieldwork had to be conducted in order to answer the research questions stated in Chapter 1. Although fieldwork is fraught with both practical and theoretical difficulties (some of which are beyond the control of the researcher), the promise of collecting *real* data more than justifies the inconvenience, uncertainty and cost associated with the endeavour.

Data collected from the field is, by its very nature, both *natural* and *grounded*, and is therefore an excellent first step in the formulation of theories which will be both useful and applicable. One of the purported strengths of fieldwork is that behaviour is observed in as natural a way as possible, such that natural sequences of events are observed in natural settings, 'on-the-spot through the eyes of a detached observer' (Brewer and Hunter 1989: 162). This

type of observation, it is argued, aids the formulation of *realistic* theories, theories which, with respect to the real world, 'work and fit' (Glaser and Strauss 1967: 3). This is an important strength of fieldwork, and one which is particularly relevant for this research, since the aims and objectives of this study are very much applied rather than theoretical (see sections 1.1 and 1.4 above).

In short, fieldwork was considered to be the only feasible way of collecting abstracts which a) were natural, b) were grounded, and c) could be reconciled linguistically with respect to certain elicited success judgements. To be natural, the abstracts had to be written by real people for real readers, and had to be surrogates for real source texts. To be grounded, the abstracts had to be collected together with data representing something of the context in which they were written. To be capable of being analysed fairly and precisely, the abstracts had to be in sufficient quantity to enable comparisons to be made between examples judged to be of different degrees of success.

Unfortunately, however, although very rich, fieldwork is arguably the hardest source from which to collect data: there are ethical dilemmas concerning how far it is felt legitimate to disrupt the ordinary everyday lives of the people being studied; although seldom discussed at any great length in the literature, there is the very real problem of gaining entry, or just 'getting in' (Becker 1970: 15 - 17), to the field of interest; there is the problem of the Observer's Paradox, together with a bewildering variety of other associated difficulties. However, many of these can be alleviated by careful planning; Chapter 6 below details the decisions that were taken to forestall some of the more obvious pitfalls of fieldwork.

In addition to these various problems, it must be acknowledged that not everything can be accomplished by means of fieldwork. In particular, the fact that fieldworkers typically have little control over what they observe often means that the data they do collect is incomplete. In fact, this was true of the fieldwork carried out for this study: while the fieldwork provided a rich supply of data in the form of texts, it was an altogether less rich supply of data in the form of judgements.

The next subsection discusses those survey methods which were used to augment the fieldwork by collecting various informants' attitudes, opinions, and reported behaviours.

2.3.2 Surveys

Chapter 7 below documents the decision-making that led to the design of the survey materials. This section will briefly summarise why *elicited* (as opposed to naturally occurring) data had to be collected in addition to the field data in order to answer the research questions stated in Chapter 1.

The naturalistic abstracts data had to be supplemented by the collection of judgements made by typical readers of such abstracts simply because the judgements indigenous to the field would not allow a scientifically satisfactory tie-up to be made between the two. In other words, standardised questions had to be put to additional informants in a more systematic fashion in order to allow hypotheses to be formulated which would account for the success judgements,.

Within the sociological paradigm there are thought to be two ways of eliciting subjects' attitudes or opinions: these can be obtained either by interviewing, or by conducting a survey (though some authorities differ over terminology, it will be assumed here that a questionnaire is a particular type of survey). As may be anticipated from earlier discussion, each technique has both strengths and weaknesses.

From the point of this study, the most relevant strength of the interview is that investigators can remain open to what their informants are telling them; in other words, unanticipated insights are not prevented from surfacing by the very nature of the inquiry. In this way interviewing is properly thought of as being a component of participant observation, or fieldwork. However, misunderstandings often arise, and interviewers should be aware of the very real possibility that they 'may interject personal biases' (Marshall and Rossman 1989: 83) into the interaction and so distort the data collected.

The most serious disadvantage of the interview with regard to the present study pertains to standardisation. While it may be possible in principle to ask interviewees all and only the questions from a strictly predetermined set, this is seldom what happens in practice, and, in any event, it is notoriously difficult to collect results which permit close and systematic comparison both between texts and informants.

Surveys, on the other hand, allow rather tighter control and in so doing promise to make 'data easy to manipulate and categorize for data analysis' (Marshall and Rossman 1989: 102) and should also lead to greater generalisation. The major disadvantage of the survey is that it runs the risk of desensitising itself to the data; Cicourel (1964: 105) warns 'some advocates of

the interview often point out that the questionnaire with fixed-choice response categories precludes the possibility of obtaining unanticipated definitions of the situation which reveal the subject's private thoughts and feelings. While fixed-choice alternatives may be adequate and necessary for obtaining factual data, seeking information on social processes by this means may force the subject to provide precise responses to events and issues about which he may be ignorant or vague'.

The multimethod approach urges that both techniques be considered, either by carrying out both types of task independently, or by combining the best parts of both into one. As Chapter 7 demonstrates more fully below, this latter option was preferred. By combining both open and fixed-choice questions, comparisons, generalisations and predictions all could be made in a principled way, while still remaining sensitive to the opinions (although quite possibly vague) which gave rise to the more immediately testable judgements. Combining the two techniques was achieved by eliciting quantitative type data in the form of rankings ('Please indicate which you thought most helpful'), together with associated qualitative type data ('Can you say why you thought this most helpful?'); Chapter 7 below provides full details.

In conclusion to this brief discussion of interviews and surveys, two further points will be made concerning their relation to the multimethod approach. In contrast to what has gone before, the first concerns an advantage they both share, the second a disadvantage.

Both techniques emphasise the importance of collecting the opinions of those people who form a sample of the larger population the investigator is interested in. When the investigator does not belong to this larger population, it is all the more important to study and value the insiders' beliefs. In so doing, the investigator treats informants as *de facto* observers (Denzin 1970: 202), itself an act of triangulation, a necessary though not sufficient condition for a more rounded investigation.

Both techniques, however, share a common failing, one which is remarkably difficult to remedy. Once again this stems from the Observer's Paradox. In giving their opinions informants are themselves being observed, and so the investigator should assume that they will react to some degree to such questioning. Nonreactive, or unobtrusive, methods have been suggested as a means of combating such a difficulty, and so the next section will assess their usefulness and applicability.

2.3.3 Nonreactive Methods

Researchers use nonreactive methods when they suspect that their presence may have an influence on that which they are trying to observe. Such methods are termed 'nonreactive', since subjects are assumed not to react to the presence of observers when the observers are completely removed from the situation, or are otherwise carrying out their study in an entirely unobtrusive way.

Suppose a somewhat ungenerous company boss suspects that his employees are drinking ever increasing amounts of tea during their breaks, so much so that they have correspondingly less time to do their work. Rather than confronting the employees and asking them about their tea-drinking habits directly, the boss would do better to monitor the number of tea-bags used from week to week, only checking the caddy once everyone had gone home.

This is a rather trivial example of a nonreactive measurement technique, but it does demonstrate how remaining completely in the background can sometimes help gain a more accurate picture of events. The kinds of nonreactive data used in more likely scenarios include 'actuarial records, political and judicial records, records of government agencies, and data from the mass media' (Denzin 1970: 265). Chapter 5 presents a review of various standards and guidelines published by the Information and Library Science communities pertaining to the writing of abstracts. These are potentially very valuable documents since one would assume that such prescriptive information should provide a great many interesting insights into what it is to write 'good abstracts'. But, although such data is technically nonreactive, it must be remembered that it may still contain bias. Thus, Denzin (1970: 265) warns 'analysts must realize that any public archival document represents the imprint of the organization that produced it, and thus bias arises simultaneously from both the author and from his organization'. Notwithstanding this admonishment, such manuals have been scrutinised to test whether or not following their suggestions can lead to the production of better quality abstracts.

In fact, the approach taken in this thesis is to view literature reviews *more generally* as being sources of non-reactive data, as well as having the more usual functions (see the Introduction to Part Two). Thus Chapters 3, 4 and 5 all shed some light on abstracts and abstracting, but each from a slightly different perspective. In addition to the normative data discussed in Chapter 5, Chapter 3 considers the view from linguistics and psychology (assumed to be

the most familiar territory), while Chapter 4 assesses work in Artificial Intelligence. In this last connection Phillips has argued (1985: 354 - 61) that the simulation of behaviour by means of some model constitutes a supplementary means of collecting data. At least some of the programs that reduce original material are claimed to be computer simulations of human abstracting ability, and so these are discussed at some length below.

By triangulating nonreactive data from different research paradigms, it is hoped that a greater number of findings may emerge that will be relevant to the fundamental aims of the research.

2.4 Data Analysis

This section shows how the data analysis phase of the research, description of which forms the bulk of this thesis, was designed. Here two features of the design will be discussed, both of which follow from the adoption of the multimethod approach: triangulation of analysis; and supplementary experimentation.

The fundamental goal of the linguistic analyses carried out in the present work is to discover what it is that is in the abstracts that gives rise to the various judges' opinions. Pretheoretically, 'success' is a multi-faceted notion, and so it is to be expected that there will be a great many different linguistic factors which can affect how successful a particular abstract is perceived to be. Various decisions have to be made, therefore, in order to narrow the scope of the inquiry to something that is manageable given the constraints of the study. Clearly it is desirable that this narrowing be done in as principled a way as is possible, so pilot analysis is used to identify the most promising avenues for larger scale investigation. Part Four below will show how different types of linguistic analysis were identified, motivated in some cases directly by the qualitative opinions of the judges, in others by the experience of other researchers engaged in similar projects.

Part Four contains some chapters which can be regarded as smaller pilot studies (in Chapters 9 and 11, only a small amount of data is used to evaluate the usefulness and appropriacy of certain types of linguistic description) and some which can be regarded as larger-scale investigations (in Chapters 10, 12 and 13, all seventeen Information Science abstracts are subjected to analysis and hypotheses are investigated). The decision to combine some of the characteristics of a pilot study with some of the characteristics of a larger-

scale study is discussed in the Introduction to Part Four below.

The aim of this phase of the research is to be able to claim, through the formal statement of hypotheses, that causal relationships hold between certain configurations of linguistic features and certain patterns discernible in the judgements. However, there is always going to be a danger in an exploratory study such as this that assumed correspondences between variables and judgements in fact take place merely by chance, or that variables become confounded such that it is no longer possible to decide what is the effect of any one variable. Some of these problems can be addressed through the careful use of controlled experimentation. Experimentation provides a precise means of assessing the validity of claims, or, more accurately, of attempting to falsify hypotheses. This is one occasion where the constraint to use only naturally occurring texts should be relaxed such that variables can be controlled. Under such circumstances the many different linguistic factors which could affect the success of an abstract can be held constant while others are deliberately manipulated. New sets of judgements can then be collected for the abstracts that have been rewritten by the analyst, and correlated against those predicted by the hypotheses.

The artificiality of this data runs counter to the predominantly naturalistic rationale underlying the design of this study, but it is nevertheless a useful complement in that it provides a powerful means of evaluating the results of the analyses to see if they have more general application. Experimentation is beyond the scope of this thesis, but further work of a more experimental nature is suggested in Chapter 15 below.

2.5 Conclusions

The criteria by which the trustworthiness of a study can be established will depend on the type of research being conducted (by 'trustworthiness' is meant the degree to which the research design provides satisfying and satisfactory answers to the fundamental questions). Traditionally, positivist studies are gauged against four criteria - internal validity, external validity, reliability, and objectivity - whereas naturalistic studies are gauged against credibility, transferability, dependability, and confirmability. This latter set of terms is derived from Lincoln and Guba (1985: 289 - 331) who discuss the interrelationship of these various measures. What still remains unclear, however, is how to evaluate an approach which borrows methodological facets

from both positivist and naturalist modes of enquiry.

This is partly due to the fact that such combining of techniques happens relatively infrequently. But this is not the whole problem; it is possible to combine styles in many different ways, each quite possibly demanding different criteria. Bearing these caveats in mind, only the first two members of each set of criteria will be considered here, since these reflect a clear overlap between the two standpoints; it is less obvious, for example, how reliability, in its positivist guise of stability, tallies with its naturalist counterpart dependability, the construct 'in which the researcher attempts to account for changing conditions in the phenomenon chosen for study' (Marshall and Rossman 1989: 146).

'Credibility' is the naturalist analogue of internal validity, the validity 'with which we infer that a relationship between two variables is causal or that the absence of a relationship implies the absence of a cause' (Cook and Campbell 1979: 37). More generally, it refers to the confidence with which findings are presented, and interpretations are derived. Spurious explanations have been guarded against in this study by using real data, viewed and analysed from a number of different perspectives. These include: 1) triangulation of *data* - as is detailed more fully in Part Three below, many different abstracts have been collected, from different writers, summarising different source texts; 2) triangulation of *method* - namely participant observation, surveys, and nonreactive methods; 3) *investigator* triangulation - in a study such as this it was not thought possible to arrange for a team of investigators to undertake the fieldwork, but this was approximated to by arranging for a number of non-field-participants to act as judges; and 4) *theory* triangulation, which allows different types of linguistic analysis to be used.

External validity and its naturalist twin transferability concern the degree to which the findings from one particular study are expected to be able to account for other settings more generally. Because naturalists believe that the universe is constantly changing, and that a particular setting can never truly be replicated, technically they are unable to specify how externally valid their enquiries are. However, in practice, they can and should tailor their work in such a way so as to enable comparisons to be made quickly and easily. The view taken in this thesis is that investigators can aid transferability by i) discussing the settings, methods and analyses of the study as comprehensively as possible (Parts One, Two and Three), and ii) by stating hypotheses and results clearly, so that these can be applied to new contexts (Parts Four and Five).

Given that the goals of this work as outlined in the last chapter are very much geared to explaining real linguistic phenomena, combining data from fieldwork, surveys and nonreactive sources should ensure that the words and judgements analysed here form a realistic sample of real-world behaviour. The desire to ask 'real questions' will not suffice, however. To provide 'real answers' to such questions, observational sophistication must be augmented by theoretical precision, such that hypotheses can be clearly derived, stated and tested. Here quantitative analysis, and close attention to analytical detail come more clearly to the fore. This pairing of research styles should help results emerge, results which may prove insightful in real field situations.

Part Two

Reviews of the Literature

Introduction

The last chapter outlined the general design of the study and discussed why it was considered necessary to adopt a multimethod style of research.

The multimethod style is evident in Part Two in two ways: firstly, it contains not one literature review but three, each pertaining to a different discipline; and secondly, each review is expected to serve more than one purpose.

Marshall and Rossman (1989: 34 - 35) state:

'The literature review serves four broad functions. First, it demonstrates the underlying assumptions behind the general research questions. If possible, it should display the research paradigm that undergirds the study and describe the assumptions and values the researcher brings to the study enterprise. Second, it demonstrates that the researcher is thoroughly knowledgeable about related research and the intellectual traditions that surround and support the study. Third, it shows that the researcher has identified some gaps in previous research and that the proposed study will fill a demonstrated need. Finally, the review refines and redefines the research questions and related tentative hypotheses by embedding those questions in larger empirical traditions.'

With reference to the first of these four functions, this thesis is an application of discourse theory; as such, the intellectual discipline which principally 'undergirds' it is a particular kind of linguistics. Work in linguistics relevant to abstracts and abstracting is therefore reviewed in Chapter 3. In addition, relevant studies from two branches of the neighbouring discipline of psychology are discussed. This chapter should serve to orient the reader to the assumptions and values brought to the study by the researcher; some of these are also implicit in the propositions stated in Chapter 1 above.

The desire to have a 'multi-review' stemmed from the belief that insights could be gained from surrounding disciplines and that the project would have suffered had it been restricted to territory most familiar to the researcher. Reviews of related work are therefore provided in Chapters 4 and 5, the first of

which describes the view of abstracts and abstracting from computer science, and more especially from Artificial Intelligence, the second that from Information and Library Science.

It will be seen that all three chapters, although interesting, fall somewhat short of providing wholly satisfactory answers to the basic research questions posed in Chapter 1 above. These are the gaps referred to in Marshall and Rossman's third function: this thesis constitutes an attempt to plug some of these gaps by laying a firmer foundation for the design of teaching materials based upon a clearer understanding of what makes some abstracts appear more successful than others.

Fourthly, it is hoped that Part Two will provide a wider context for the basic research questions of the study. Although it is not explicitly argued, it is thought that the type of methodology adopted in this research could be used to advantage for the investigation of other types of problem found in the three disciplines.

In addition to the four purposes Marshall and Rossman describe, it is hoped that Part Two will fulfil a fifth: in section 2.3.3 above it was stated that the literature reviews would also act as a source of nonreactive data. This is especially true of Chapter 5, which reviews a selection of published guidelines and standards for the practice of abstracting, the assumption being that adherence to these prescriptions will result in better quality abstracts.

3

Review of the Psychology and Linguistics Literature

3.1 Introduction

This chapter reviews work concerning abstracting and summarising from three different research paradigms. Section 3.2 provides some brief comments on Cognitive Psychology, but, since much of this work is of only peripheral relevance to the basic research questions of the present study, the reader is referred to a more comprehensive overview of the tools and techniques available (Gibson 1989). Section 3.3 details two Experimental Psychology accounts of the processes by which information becomes transformed from original to summary versions. Finally, section 3.4 reviews work from Linguistics relevant to abstracting.

3.2 Some Brief Remarks Concerning Cognitive Psychology

Cognitive Psychology looks extremely promising for the study of abstracts and abstracting because notions such as 'summary' and 'gist' are investigated more extensively than in any other paradigm. However, much of the work is characterised by a separation between higher-level knowledge structures and their underlying linguistic realisations, the emphasis being very much on the former. In fact, relatively little attention is paid to actual words, and naturally occurring texts are seldom considered. This makes much of the research of only limited practical utility for discourse analysis (c.f. Brown and Yule's comments, 1983: 114 f.f.).

Gibson 1989 provides an overview of the different methods of text representation which form the basis of the various cognitive theories of

summarisation. Three methods are evaluated: super-structures (see for example Kintsch and van Dijk's various propositional theories of text comprehension: Kintsch 1974, Kintsch and van Dijk 1978, van Dijk and Kintsch 1983); parse trees (see for example the work on story grammars: Rumelhart 1975, 1977, Mandler and Johnson 1977, Thorndyke 1977, Stein and Glenn 1979 and Peterson and McCabe 1983); and hierarchies of rhetorical predicates (see McKeown 1982, the work of Mann and co-workers, for example Mann 1984, and Fox 1987).

The next section discusses work conducted in Experimental Psychology which is more empirical. Greater attention is paid to the performance of subjects summarising texts; also, their summaries are analysed to a level of detail beyond what is typically attempted in Cognitive Psychology.

3.3 Review of Work Relevant to Summarisation in Experimental Psychology

There is much work in Experimental Psychology which investigates summarising skills. However, in this brief review none of the work which assumes summarisation to be a measurable intellectual yardstick and which seeks to investigate how various conditions can affect this skill is considered. These studies include accounts of hearing-impaired subjects (Akamatsu 1988; Peterson and French 1988), handicapped subjects (Rie and Yeh 1983; Reis and Leone 1987), and alcoholics (Turner and Parsons 1988; Schaeffer et al 1988). Neither are those studies considered which centre on the relationship of summarising to: recall (Spurlin et al 1988; Munch and Swasy 1988); reading (Walker 1985); nor to studying more generally (King et al 1984; Rinehart et al 1986).

The present thesis makes no attempt to reconcile the abstracts with the original source texts. This decision was taken so as to limit the scope of the investigation, and because abstracts are seen by readers to be texts in their own right. Studies of abstracting in Experimental Psychology, on the other hand, typically deal with the relationship which holds between the summary and the original text.

Two strands of research are reviewed in what follows. The first (section 3.3.1) is essentially atheoretical, and maps correspondences between original and summary (Sherrard 1986). A small amount of data from the present study is briefly examined to test Sherrard's claims. Section 3.3.2 reviews Johns 1985

which builds on the mapping idea introduced in the previous section. Also surveyed in Gibson 1989 are a number of articles by Brown and her colleagues (see for example Brown and Day 1983 and Brown et al 1983) which make much more explicit use of insights borrowed from van Dijk's macro-processing approach, together with work by Armbruster et al (1987) which demonstrates how instruction in problem-solution schemata can facilitate summary writing.

3.3.1 Sherrard 1986

Sherrard 1986 plots correspondences between sentences of source texts (sentences being defined orthographically) and those of their summaries, and then examines the patterns which emerge. Her own word for this process of mapping the various linkages is 'topography', a term which properly suggests that it is predominantly the surface features of the texts which are the objects of experimental scrutiny.

In this way the article examines what (subset of the) information from the original appears in the summary, but does not discuss how exactly it is realised linguistically. Her major results concern the types of mapping encountered and their various frequencies. The most favoured translation between original and summary was one-to-one, and accounted for 40% of all summary sentences (1986: 331). That is to say, a summary sentence thought to have its origin in one sentence of the original only would be considered to be a one-to-one mapping, diagrammed as follows:

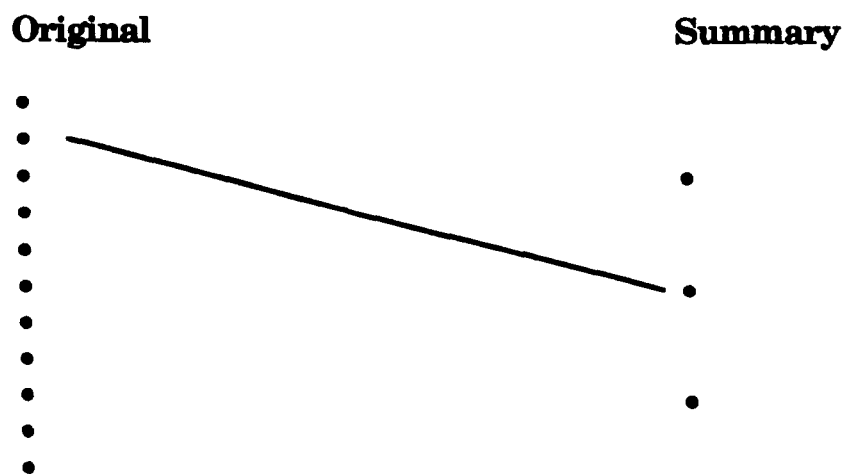


Figure 3.1: a one-to-one mapping (after Sherrard 1986).

The next most frequent strategy was two-to-one mapping, diagrammed as in Figure 3.2, followed by 'triples', and then 'quadruples'.

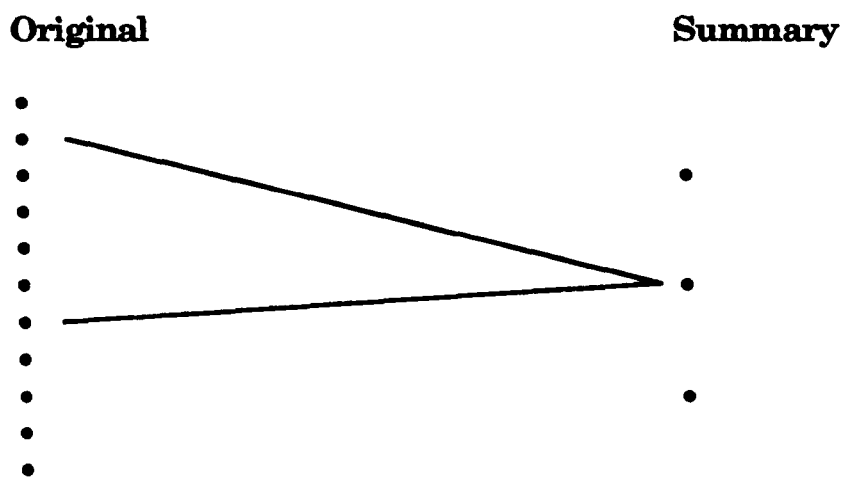


Figure 3.2: a two-to-one mapping (after Sherrard 1986).

Simple one-to-one selections and simple deletion (i.e. disregarding, or choosing not to represent, original material) are together claimed to account for 82.5% of all text sentence - summary sentence mappings in her corpus of 7 originals and 33 summaries (1986: 331), a surprisingly high percentage.

As a brief test of this claim, an abstract was examined from the present study which summarised the article on the 'Brain Drain' from the Independent newspaper. The first sentence of Abstract C is reproduced here:

Brief article predicting that with the fall in student numbers in 1992, owing to the decline in 18-year-olds entering higher education, employers will have to use more imaginative and flexible recruitment techniques in order to attract graduates.

(Sentence number 1, Brain Drain, Abstract Version C).

According to the present researcher's intuitions, this sentence appears to derive from (at least) the first three sentences of the original, the following list of words or phrases being directly mapped from original to summary:

fall (from original sentence 2), student numbers (from 2), in 1992 (from 2), the

... decline in ... 18-year-olds entering higher education (a selection taken from the clause-initial nominal group of 1), employers (from 1 and 3), graduates (from 3).

There are a number of additional 'weaker' links, if words are allowed to become re-classed grammatically. So, whereas 'fall' appears identically in both original and summary, in spite of its being transformed from verb to noun to form a larger nominal group, 'imagination' reappears in adjectival form, and the 'recruit' verb of the original becomes nominalised as 'recruitment'. If such transformations are allowed to constitute legitimate linkages, there then remain the following emboldened sections of sentence 1 of Abstract C still unaccounted for:

Brief article predicting that with the fall in student numbers in 1992, owing to the decline in 18-year-olds entering higher education, employers will have to use more imaginative and flexible recruitment techniques in order to attract graduates.

These emboldened sections are new to the abstract. 'Predicting that' is illocutionary metadiscourse (see Stainton forthcoming (b)), while the 'owing to' is included such that the decline can be signalled as being causative in a dependent clause; this contrasts with the more congruent realisation in the main clause of the first sentence of the original. In both cases, however, new constructions are generated by the producer of the abstract, rather than the original simply being edited in some way.

Although a study of the relationship between the source text and its abstract lies beyond the scope of this thesis, it would be interesting to hypothesise that more successful abstracts would contain significantly greater proportions of 'new' material than would unsuccessful abstracts.

3.3.2 Johns 1985

Some of Sherrard's hypotheses have been independently investigated, albeit in slightly different form, by Johns 1985; this article examines the relationship between original and summary in generally greater analytical detail, but has relatively little to say concerning the topographic linkage between source and condensation.

Johns points out that the means by which university students summarise real texts still remain relatively unexplored. Her own work is an attempt to redress this imbalance, and has a primarily educational motivation. Teaching students how to write quality summaries, she argues (1985: 497), requires 'a comprehensive scheme from which completed summaries can be coded to indicate how students reduce, replicate and distort the original text ... in order to determine what these students' skills are, and what needs to be done to improve the teaching of summary writing at this level'.

Johns' work goes beyond Sherrard's topographic study in that Johns does not merely chart correspondences quantitatively, she also attempts a qualitative assessment. Implicit throughout the article is the idea that there is a right and a wrong way to represent important information; her own phrase refers to the way in which information is either 'replicated or distorted'. She claims then that poorer, or 'underprepared', students are more likely to distort the meaning of the original, and that this will have a negative effect on the perceived quality of their work.

Johns postulates that a summary 'idea unit' ('IU': Kroll 1977; roughly corresponding to what Halliday 1985 defines as a clause) is either a correct replica or some kind of distortion of some unit (or combination of units) from the original. Reproductions correspond to Sherrard's one-to-one mappings, and can either be direct copies, or paraphrases of single original IU's. Combinations are many-to-one type mappings, and are classified either as being combinations across, or within, paragraphs. Johns departs from Sherrard, however, with the proposal of a third type of replication, macro-propositions, as the 'accurate, writer-invented statements which provide the "gist" of a paragraph or the reading' (1985: 515). She posits three types of macro-proposition: those conveying the meaning, or gist, of a paragraph; those conveying the meaning, or gist, of the entire text; and what she calls 'meta-statements', an example being 'In this section, "Anti-War Sentiment in the North", the author lists reasons for the sentiment of the people during the war' (1985: 515), a particular type of informational metadiscourse (see Stainton forthcoming (b)), somewhere between a pre- and a post-plan.

Figure 3.3 below illustrates Johns' classificatory framework for each summary idea unit:

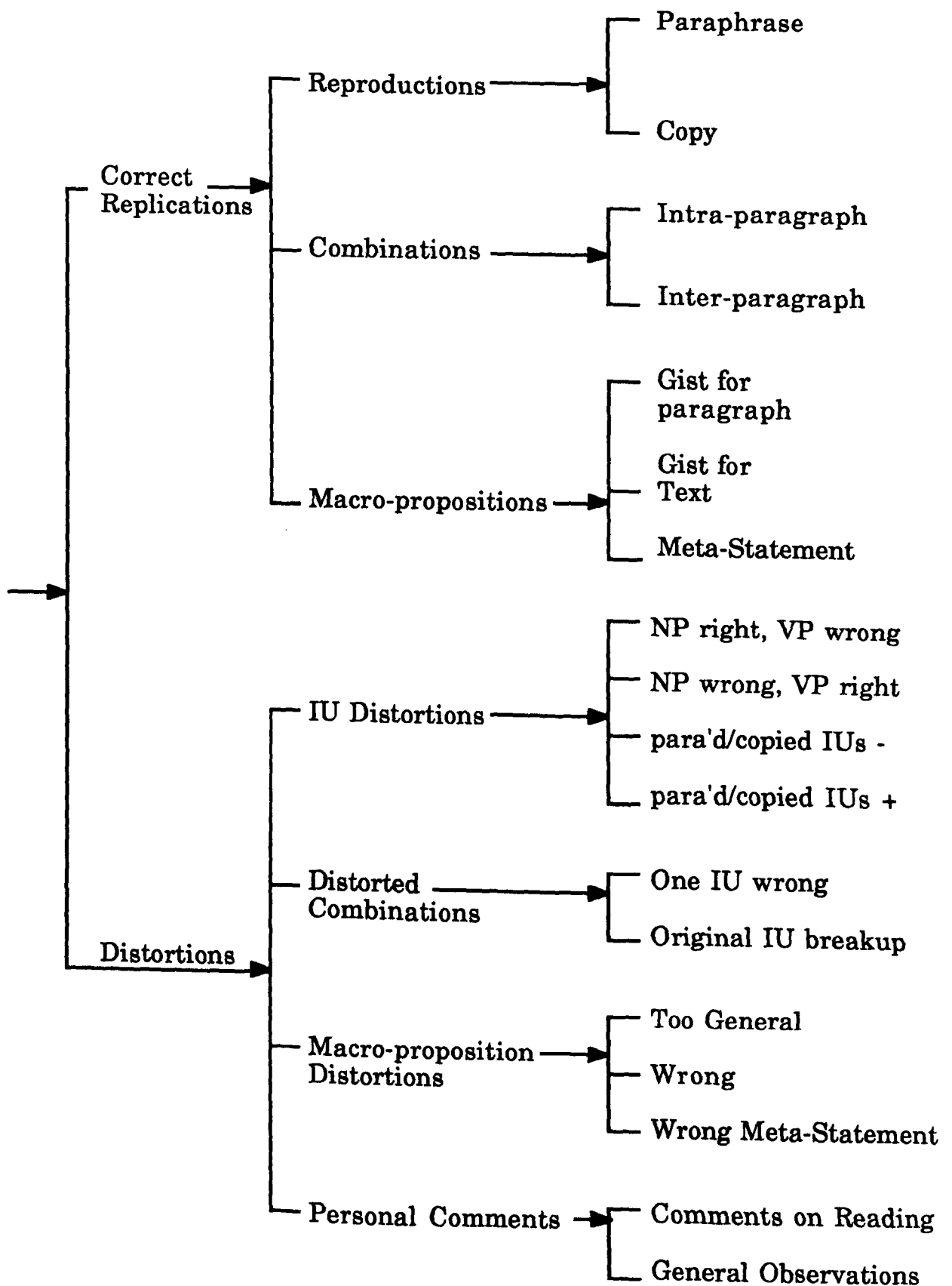


Figure 3.3: After Appendix B, Johns 1985: 515 - 517

Turning now to those strategies which it is assumed will make for a poorer

quality summary, the distortions are subcategorised as follows. Firstly, there are those IUs from the original which become in some way deviant. These are one-to-one mappings which either become in some way extended or pruned, or some selection of which becomes corrupted. In this way, Johns suggests that summary IUs can faithfully represent the NP of an original IU, but can misrepresent its VP. Alternatively, the VP can be satisfactorily expressed, but with a deviant NP. These can both be thought of as corrupt transformations between original and summary. The remaining types of IU distortions are additions and deletions. The second major type of distortion are those which involve the mis-combining of original IUs into a single summary IU. There are two types: two or more combined summary IUs, one of which is inaccurate; and 'breakup of combinations of idea units in the original with no additional information included [in the summary]' (1985: 516). Thirdly, macro-propositions can backfire if they present information which is overly general, inaccurate, or if they result in unfelicitous metadiscourse. Lastly, personal comments seem to be generally frowned upon, and would include comments on the reading, such as 'It seems as though everyone was against Lincoln' (1985: 517), and more general observations, such as 'This was a boring reading, because all history readings are boring' (1985: 517).

If it could be made more precise linguistically, Johns' framework may provide a useful means of characterising the success of a summary primarily in terms of its content. However, as will be demonstrated below in Chapter 8, many of the reasons which judges believe underlie their preferences go beyond the raw information expressed in the abstracts and concern the way in which the information is realised. Issues of readability, for example, are particularly important; it is not sufficient merely to represent pertinent information, it must also be expressed clearly and coherently.

3.4 Review of Work Relevant to Abstracting in Linguistics

Unlike Experimental Psychology, there is relatively little work in textlinguistics which addresses the problems of abstracts and abstracting.

Two strands of research are reviewed in what follows. The first (Nwogu and Bloor 1989; section 3.4.1) is representative of a small body of work carried out within the framework of Prague School linguistics, and uses Thematic

Progression as its basic analytical tool. Section 3.4.2 reports on a study carried out from a Systemic-Functional perspective (Drury 1989) and is directly relevant to the aims and objectives of the present study.

3.4.1 Nwogu and Bloor 1989

Nwogu and Bloor discuss three different types of text. These all have a medical theme, but each discusses findings, implications, and so on, in a slightly different way. The first type is the 'research article proper' (or 'RAP originals', 1989: 1), exemplified by articles in publications such as the *British Medical Journal* and the *Lancet*. The second type is the accompanying abstracts (ABS) of the original RAPs; Nwogu and Bloor view these two types as belonging to the professional genre of written medical discourse, while their third type is regarded as more popular, or journalistic. The Journalistic Report Version (JRV), therefore, presents the same essential discourse contents as the RAP, but in a manner which is more accessible to lay-people; the JRV originals are taken from the *New Scientist*, *The Times*, and *Newsweek*. Their corpus consists of 45 texts: 15 RAPs; 15 ABSs of these RAPs; and 15 JRVs.

Nwogu and Bloor's basic hypothesis is that some of the contextual differences between the three types will give rise to differences in their thematic structure. Like Stainton (forthcoming (c)), Nwogu and Bloor investigate theme according to two descriptive frameworks: they analyse their texts for thematic progression ('TP'; see for example Danes 1974, Firbas 1987), and, in addition, using Hallidayan grammar, they discuss the elements which realise theme. However, whereas Stainton believes Halliday's approach to be more useful, Nwogu and Bloor devote more space to the intricacies of TP.

Simple Linear and Constant TP occur frequently in all three types, while the ABS 'makes fairly even use of both patterns' (1989: 6). Simple Linear is found more often in the JRV than the others, whereas the RAP makes heaviest use of Constant TP. The authors tentatively suggest that this may be due to the relative degrees of shared knowledge between author and (the author's perception of) the reader, such that Simple Linear could be thought of as being 'a means of improving on the amount of mutual knowledge' and a 'means for ensuring that the reader's involvement in the text is sustained, for, by continuously providing the reader with given information which was the previous utterance's new information, the author engages the reader in a

dynamic process more stimulating than that of the Constant Theme with its more plodding movement' (1989: 7).

Notice that the basic opposition implicit throughout the article is that between the professional and popular genres, so the similarities and differences between the two professional genres, the RAP and the ABS, are only touched upon briefly. Interestingly, Nwogu and Bloor find no occurrences of the derived pattern, nor the split rheme, in any of the ABSs (1989: 9).

Turning now to theme type, the basic difference between the RAP and the JRV seems to be that the JRV favours the representation of human agents, together with suitable pro-terms. So, whereas the RAP features:

- (1) simple nominal group with a common noun as head;
- (2) complex nominal group with a common noun as head;
- (3) prepositional phrase (adjunct) plus nominal group
with a common noun as head;
- (4) adverbial group (adjunct) plus nominal group
with a common noun as head;
- (5) empty existential there (1989: 10),

the JRV prefers elements (1989: 11) such as:

- (1) proper nouns (or nominal groups - sometimes compound - with proper nouns as head or in apposition or as modifier);
- (2) personal pronouns;
- (3) demonstrative pronouns;
- (4) adverbial group (adjunct) plus nominal group
with common or proper noun as head.

This is not entirely surprising, Nwogu and Bloor argue, since one would expect the RAP to emphasise the research over its instigators. They imply (1989: 10) that such depersonalisation is also a feature of the ABS, but do not investigate further. This difference is also apparent in the Information Science abstracts examined in the present study: it may be hypothesised that informative abstracts (see Chapter 5 below) will behave more like the RAPs, whereas indicative abstracts will behave more like JRVs. This is because informative abstracts typically seek to convey as much of the salient content of the original as is practical, whereas indicatives seek merely to give a flavour of the work. These points will be investigated in Chapter 13 below.

3.4.2 Drury 1989

The study which has most in common with the aims and objectives of this thesis is Drury's account of summary writing written from within the EFL perspective (1989a, see also Drury 1989b). Since her motives are primarily educational, she is necessarily interested in the notion of success, and argues that if one's task is to teach foreign learners of English how to write successful abstracts, then the 'language forms adopted by successful writers ... need to be described' (1989a: 1). Because her analyses are often quite detailed linguistically, her text base is rather small: she discusses three summary versions of one original, Kalivoda 1980, a predominantly expository article examining the arguments concerning the best way to teach listening skills in ESL. These summaries were naturally-occurring and were in fact a component of some of the assessed work leading to a TEFL diploma at Sydney University. Three summaries were written: one by an Australian (N[ative]); one by a Botswanan (S[econd language]1); and one by a Korean (S2). Drury's subjective data, which was also naturally occurring, consists of the single grades (and associated comments) each student received from an (unknown number of) assessor(s). These are reproduced here:

- | | | |
|----|----|---|
| N | A- | weaker on relevance section, not innovative |
| S1 | B+ | good introduction, relevance section OK, did all that was required |
| S2 | B- | relevance section and/or introduction very poor, summaries generally good |

It is important to point out that these comments and ratings cannot be interpreted as being straightforward success judgements concerning solely the summaries themselves, as the marks pertain to a larger exercise in which students were asked to select a topic, to read four journal articles, and then to write an assessment of the topic to include: 1) an indication of the reasons for choosing that particular topic, and its relation to their own particular teaching situation; 2) a short summary of each article; 3) a critical evaluation of each; 4) further commentary on their relevance; and 5) a short conclusion. Clearly the

summary writing represents a component of the larger exercise, and so one cannot reasonably expect the umbrella markings to relate to the quality of the summaries in any direct way.

In what follows, Drury's analysis of theme, transitivity, and grammatical metaphor will be discussed in turn.

Firstly, relying upon the Hallidayan treatment of Theme, Drury points out that 'the N text is characterised by impersonal Themes with a high lexical density' whereas the second language students' abstracts 'tend to have more human Themes, namely, the human elements in the field - teachers and students' (1989a: 9). It is almost as if the non-native students are trying to write JRVs, while their native counterpart favours modes of expression more characteristic of the RAP.

One of the examples Drury cites to support this observation is the following 'minimal pair':

N:

3. The causes of this problem; memory, rapid and slurred speech, and unfamiliar vocabulary

S2:

3. First, students learning English as a foreign language have the difficulty of remembering

4. Second, foreign language learners have the difficulty of the rapid-sounding pace of the speech.

(Drury's underlinings are retained to denote Theme-Rheme boundaries).

This example demonstrates nicely the differences between personalised versus depersonalised thematic strategies, but also contains another point of interest. Here the second language learner is seen to enumerate her themes, thus violating Nwogu and Bloor's principle that Derived and Split Rheme TP are typically disfavoured by abstract writers. Nwogu and Bloor explain this by arguing that such a pattern is essentially an expansionist one, and therefore unlikely to find its way into what is meant to serve as a condensation of the original.

Drury finds that the different abstract producers make different use of textual themes (Halliday 1985: 54). S1 makes heaviest use, containing, in fact, 'twice as many clauses as the other summaries', N the least, with S2

somewhere in between. Again, Drury's remark that the preponderance of textual themes together with the more human thematic focus tends to place the second language learners' abstracts nearer the spoken abstract end of the speech/writing continuum (1989a: 10) tallies with Nwogu and Bloor's opposition between the professional and popular genres, and would seem to support the hypothesis that the second language learners are trying to write JRVs (and possibly even indicative, rather than informative, abstracts; see Chapter 5 below).

To give an example, S1's formulation 'He suggests that the teacher' becomes recast in N's 'Certain suggestions are made' in which the theme is depersonalised, the nominalisation making for a less congruent means of expression. The conclusion to Drury's section on theme (1989a: 10) is worth quoting in full: 'the weaknesses in the second language summaries are reflected in their Thematic choices characterised by a high proportion of human and textual elements. On the other hand the native speaker summary has successfully chosen Thematic elements which reflect the source structure and yet show a varied combination of source lexis'.

These findings are particularly interesting, because, as will be demonstrated below in Chapter 13, precisely the reverse seems to hold in the case of the Information Science abstracts: there is evidence to suggest that personalised (or what will be referred to as 'non-informational') Themes do in fact find favour with many of the judges.

Regarding transitivity, Drury notes that the L2 summaries make heavy use of mental and verbal processes, while the N is more evenly distributed. Further, 'the N text is characterised by non-projecting, mainly passive reporting structures without mediums. The S texts, however, use far more projecting reporting structures and active structures are also more common especially in S1' (1989a: 11). Transitivity is not, however, investigated in the present study because it was thought that relatively few of the judges' qualitative remarks could be explained by making appeal to an analysis of transitivity (see the Introduction to Part Four below).

Lastly, Drury considers grammatical metaphor (Halliday 1985: Chapter 10), and, following techniques presented in Ravelli 1985 and Jones 1988, provides the following measures:

	<u>N</u>	<u>S1</u>	<u>S2</u>
Grammatical Intricacy	5	11	7
Lexical Density	6.76	3.48	6.4
Grammatical Metaphor	3.4	0.86	2.56

Drury notes that statistics support the hypothesis that S1's abstract is more consonant with an oral retelling of the argument than with a written report. The 'best' summary, N, scores highest on grammatical metaphor with a concomitantly high lexical density, and low grammatical intricacy. She makes two further claims concerning metaphor: firstly, that it is an essential tool in building a summary since it 'enable[s] the reduction, generalisation and integration of information from the source' (1989a: 16); and, secondly, that its effective utilisation will be enormously difficult to teach.

Grammatical metaphor, however, is not investigated in this thesis. This is because it is still an emerging type of analysis and difficult to carry out reliably. Also, as was the case with transitivity, an analysis of grammatical metaphor does not seem to be particularly well motivated by the judges' opinions. However, the other two types of analysis are investigated below in Part Four: lexical density is examined in Chapter 10; and grammatical intricacy is examined in Chapter 12. Interestingly, both Drury's hypotheses are falsified in the case of the Information Science abstracts where the better abstracts seem to be characterised by lower levels of lexical density and by higher levels of grammatical intricacy.

3.5 Conclusions

The following conclusions are to be drawn from this chapter:

- The Cognitive Psychology literature provides some useful background information concerning the higher level processes which underpin the summarisation task. However, relatively little attention is typically paid to linguistic matters and to naturally occurring texts; this makes the literature of limited practical utility for discourse theorists.

- The Experimental Psychology literature, on the other hand, places far greater emphasis on the actual words of, in many cases, naturally occurring originals and summaries, and many interesting and promising insights are provided concerning the way information becomes transformed from source to summary.

- A small number of studies from the Linguistics literature explicitly investigate abstracts and summaries as texts in their own right using detailed, replicable linguistic analyses. Nwogu and Bloor 1989 point out similarities and differences between abstracts, their professional source texts, and more popular, journalistic accounts of the same subject matter. Drury 1989a is a particularly relevant account because she examines the relative degrees of success in summaries prepared by native and foreign writers, testing hypotheses using grammatical techniques from Systemic-Functional theory. Of these, the following will be investigated below in Part Four: lexical density (Chapter 10); grammatical intricacy (Chapter 12); and choice of Theme (Chapter 13).

4

Review of Automatic Summarising Systems

4.1 Introduction

This chapter reviews various computational systems which automatically summarise textual input. The chapter is divided into two sections. Section 4.2 describes those systems that can be considered to be 'noticers': these 'spot' statistical patterns in the text, score sentences according to centrality, extract the highest scoring sentences and concatenate these in original text order to form an extract. One example system is described in 4.2.1. Computers can perform such a cursory activity remarkably easily and efficiently, but the essentially blind nature of the search, or the fact that the machine does not 'understand' what it sees, means that the results are often less than satisfactory. At the other end of the scale lie those machines which attempt a deeper, more directed, understanding. The process is somewhat like the way people flick through newspapers until their attention is caught by a particular item. Section 4.3 describes those systems which rely on pre-defined knowledge structures to identify the noteworthy sections of the original text. One example system is described in 4.3.1. Finally, brief conclusions are provided in section 4.4.

4.2 Introduction to the Shallower 'Spotting' Systems

The following case study exemplifies the types of 'shallow' system which have been constructed to produce automatic abridgements of longer source texts. Hoey (draft MS) is used for illustrative purposes because many of its fundamental premises are shared by other systems which attempt to spot the important sentences in texts: see for example Luhn's seminal work on 'auto-

abstracts' (1958); Baxendale 1958; and Edmundson and Wyllys 1961.

The interested reader is referred to Gibson 1988, which reviews these systems, together with more recent approaches to spotting, specifically ADAM, 'Automatic Document Abstracting Method' (Pollock and Zamora 1975); and Paice's work on 'self-indicating phrases' (1981), which uses metadiscoursal items to identify the salient sections of original material.

4.2.1 An Example: Hoey's Lexical Connectivity Nets

Hoey (draft MS) argues that charting the connections between lexical items will enable any type of non-narrative text to be summarised. Unlike the systems that attempt to 'understand' their input, Hoey's approach pays no attention to the deeper psychological processes underlying discourse comprehension. So where, for example, FRUMP (see section 4.3.1 below) works by accessing pre-defined high level templates that are realised in the text under scrutiny, his interest is rather in those 'surface' aspects of the text (see Hoey 1983) which alert readers to its organisation.

The following discussion describes what Hoey refers to as a 'connectivity net'. This concept is best introduced through the use of Hoey's own metaphor, that of the academic paper. Consider the case of Smith who is a fictitious academic and has been writing one paper every year for the last few years. Hoey explains that Smith is an honest academic since he refers to his previously published papers only when there is a genuine need. If paper B refers to an earlier paper A, the two can be said to be related.

By taking account of which papers point back to, or pave the way for, other papers, one can draw lines to indicate relationships between different papers, as shown in Figure 4.1:

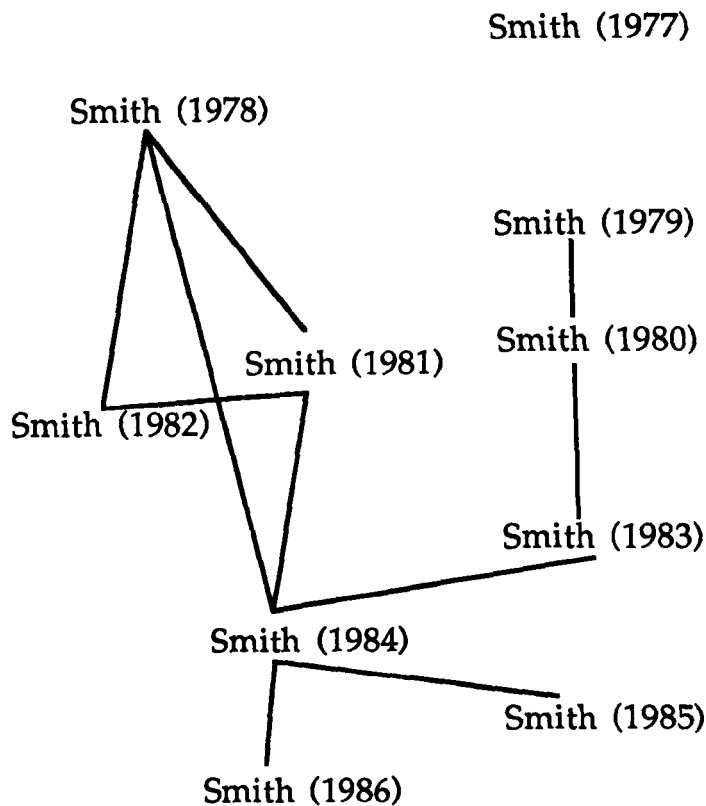


Figure 4.1: a Connectivity Net

Notice that the spacing or layout does not matter; it is the connections that are of interest. One can see from this net that Smith (1977) is out on its own; perhaps it concerns a different subject altogether. Much more representative of his work are Smith (1978, 1981), which have three ties each, and Smith (1984), which has five ties. A system of coordinates could be constructed, one of the entries for which would be:

(1978) (0,3)

This would signify that zero represents the number of works the paper refers to and that three represents the number of works that refer to it. Computing such a coordinate system provides several types of information. A high total for one particular paper would indicate that it is central to the author's thinking; a low total indicates marginality; a high first coordinate means that that paper in some sense summarises a number of previous strands of thought; a high second coordinate means that that paper may point forward to new topics investigated in subsequent studies.

The idea, then, is to do exactly the same for texts, as a theory of coherence

and cohesion will have to show how particular strings of words fit in with their neighbours in the discourse. Instead of charting connections between papers that refer to one another, some new matching criteria will have to be formulated to chart correspondences between words in a text. In Hoey's theory there exist a number of different types of relation that may hold between words in a text, the most basic of which is thought to be that of repetition. The types of relation include the following:

- (1) Simple repetition. Exactly the same word in two different sentences counts as simple repetition, but minor changes to the word also come under this heading;
- (2) Complex repetition. 'dancing' recurring as '[to] dance', for example, is thought to be complex since there is a change of grammatical class. 'fool' repeated as 'foolishness' fits in here (though there does appear to be a cline between 1 and 2);
- (3) Substitution, including elements such as pronouns, discourse deictics, demonstrative adverbs and the like (for an explanation of these terms, see Quirk et al 1972);
- (4) Ellipsis;
- (5) Paraphrase. Ideally, it would be advantageous to provide the system with an online thesaurus. For example, the program should know that, say, 'wage' (noun) is semantically linked via the concept of paraphrase to 'pay' (for the sake of argument, verb);
- (6) Superordinate and hyponymic relations. Included under this category are relations of the scientist - biologist (superset - subset) type.

Once the connectivity net has been generated, which forms a web of links between the different words in a text, each word's centrality score is calculated, and those sentences which contain the greatest amount of central words are extracted and concatenated in original text order to form the abridgement of the longer source text.

To summarise, the function of a connectivity net is:

- a) to reflect the relationships between different parts of a text;
- b) to track the development of a theme, or thread, running through a text;
- c) to establish topic boundaries in a text; and
- d) to provide a centrality/marginality rating for all sentences in a text.

The advantages of such a spotting system are as follows. Firstly, it is relatively computationally inexpensive. Secondly, the system can take naturally occurring discourse as its input. Thirdly, the summariser is unrestricted, in that texts can be on any subject, provided they are not narratives.

The disadvantages are as follows. Firstly, provision of a thesaurus is a complex matter, but one which will very probably improve the performance of the system considerably. Secondly, spurious connections will be generated at least some of the time. False ties are an inevitable feature of the surface type of approach. Thirdly, the notion of what constitutes a successful summary remains ill understood. This last point is an important one. Studies that systematically aim to point out the similarities and differences between summaries that are extractions from the source text and summaries that are person-generated are required to enable principled improvements.

4.3 Introduction to the Deeper 'Understanding' Systems

The following case study from computational linguistics exemplifies the types of 'deeper' system which have been constructed to produce automatic summaries of longer source texts. FRUMP (DeJong 1982) is used for illustrative purposes because many of its fundamental premises are shared by other systems which attempt to understand texts: see for example MARGIE (Schank, Goldman, Rieger and Riesbeck 1973 and 1975); SAM (Cullingford 1978); PAM (Wilensky 1978); CYRUS, a program that could answer questions concerning the professional life of former American Secretary of State, Cyrus Vance (Kolodner 1980); QUALM, a question answerer extending the scope and performance of SAM-type programs (Lehnert 1980); McMAP, an understander driven by specific domain-specific tasks: telling someone road directions, for example (Riesbeck 1982); SCRABBLE, software which summarised unexpected information (Tait 1982); BORIS, a system to comprehend narratives (Dyer 1983), and so on.

The interested reader is referred to Gibson 1988, which reviews many of these systems, together with more recent approaches to text summarisation, specifically SUSY, a SUMmarising SYstem (Fum, Guida and Tasso 1984, 1985a, 1985b), and TOPIC, 'Text-Oriented Procedures for Information management and Condensation of expository texts' (Hahn 1984, Hahn and Reimer 1984, Kuhlen 1983a, 1983b, Reimer 1984, and Reimer and Hahn 1983).

4.3.1 An Example: FRUMP

The rationale of FRUMP derives from the following statement: 'people often do not even read every word of every sentence they see. They can, often without significant loss of meaning, skip words or phrases, fail to finish sentences, or even skip whole paragraphs because of strong contextual predictions that make that extra work seem unnecessary or because they lack interest in what they expect to see' (Schank and Burstein 1985: 154). FRUMP was designed to skim and summarise naturally occurring news articles on a wide variety of topics and its inventor claims (DeJong 1982: 149) that it 'routinely produces correct summaries for stories for which it has not been primed'. FRUMP's input was restricted to texts organised in narrative fashion and the system was endowed with sufficient world knowledge, rather than strictly linguistic knowledge (no real grammatical modelling is attempted, for example), to 'understand' its input.

FRUMP differed from earlier systems such as SAM (Cullingford 1978) in that it was even more predictive, and it proceeded through recourse not to scripts (Schank and Abelson 1977; also Schank and Riesbeck 1981), but to 'sketchy scripts' (\$\$'s), which were less detailed and contained only the important events (DeJong 1982: 150). FRUMP 'knew' 60 different sketchy scripts, one of which was \$\$DEMONSTRATION. In such a \$\$, there would be a number of predicted events encoded in conceptual dependency form ('CD': Schank 1972, Harris 1985, Grishman 1986; see also Gibson 1988) but reproduced here 'in English' for reasons of clarity:

\$\$DEMONSTRATION

predicted event 1: demonstrators arrive at demo location

predicted event 2: demonstrators march

predicted event 3: police arrive

etc.

Such information is thought to constitute the 'context' necessary to understand texts. Different domains can be handled by the provision of more context, in the form of additional \$\$'s.

FRUMP works as follows. Faced with a novel text, FRUMP has to divine which of its knowledge templates to apply. This is done in bottom-up fashion typically by scanning for key words or phrases that are central to any particular \$\$ (s). Additionally, a whole event might activate a \$\$, such that the reporting of police apprehending a suspect would awaken \$\$ARREST.

Once a \$\$ has been activated, FRUMP's two principal organising modules, the PREDICTOR and the SUBSTANTIATOR, come into play. These are shown in Figure 4.2 below.

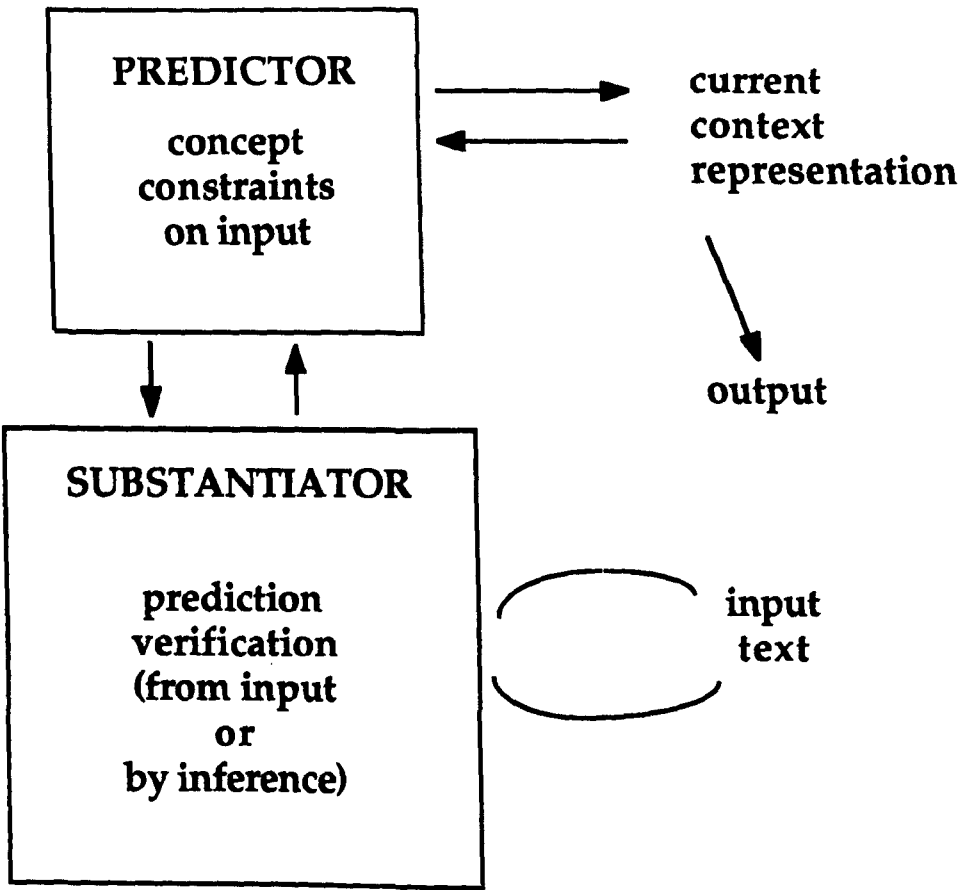


Figure 4.2: the Structure of FRUMP

The PREDICTOR considers the \$\$ to see what is likely to occur next; for example, it might state that 'the actor of the next input event will be either a head of state or a military spokesperson' (DeJong 1982: 157). In general, it attempts to make predictions on the basis of what it knows already (its current

context), the context growing all the time as more input gets processed. The SUBSTANTIATOR attempts to match the predictions against the incoming text, all this being done at the CD level. If the SUBSTANTIATOR can verify one of the predictions against one of the CD forms it has computed for a particular input sentence, it informs the PREDICTOR of the relevant CD, and the PREDICTOR in turn makes use of that CD to redefine its predictions. The constant redefining may take the form of generating totally new predictions, or retracting, or modifying, old ones. In this way only the particular phrases needed for the instantiation of certain slots of the \$\$ are searched for. Those remaining sections of the input text are deemed uninteresting and are ignored.

As with many exponents of the scripts methodology, the lexicon involved must provide a lot of information for each word, since this is the principal means of \$\$ selection. In cases where not enough knowledge had been predefined (or perhaps when that knowledge had not been specified carefully enough), the wrong \$\$ was sometimes selected. This, coupled with the fact that the system only makes an incomplete reading anyway, and therefore is only receiving a small proportion of the textual cues, meant that FRUMP only managed to read about 20% of the stories it was subjected to. One of the classic errors that FRUMP made was to summarise the headline 'Pope's death shakes western hemisphere' as 'There was an earthquake in the western hemisphere. The Pope died', owing to the misapplication of \$\$EARTHQUAKE.

FRUMP was a pioneering system which had much to recommend it. However, its input was restricted to narrative texts. Further, it was restricted to certain types of narrative text; FRUMP could not successfully process stories which it had not been told to expect. Part of the reason why FRUMP misconstrued even predictable narratives is that it has very little properly linguistic knowledge; much more importance is attributed to higher level issues, such as memory and the representation of knowledge.

4.4 Conclusions

The following conclusions are to be drawn from this chapter:

- All computational summarising systems suffer from some combination of the following three weaknesses: naturally occurring texts are seldom considered; insufficient attention is paid to the way humans write abstracts or summaries; further, the quality of the output has never been properly evaluated by asking readers whether they think the resulting abstracts or summaries are successful.

- With regard to the abstracting process, Kuhlen, one of the inventors of TOPIC (section 4.3 above), concludes 'Cognitive psychology and artificial intelligenced [sic] have, so far, not provided us with sufficient knowledge about the processes really going on in abstracters' minds when they understand texts and condense them. Thus the direct imitation of an intellectual procedure such as abstracting seems to be out of reach' (1983a: 14 - 15). An attempt is made to collect intuitions concerning the abstracting process in the present study; see Part Three below.

- Perhaps even more seriously, there have been no rigorous attempts to compare and contrast abstracts generated by computer and abstracts produced by humans. Further, computer-generated abstracts are rarely presented to real abstract readers to evaluate.

- Ideally, if computers are to be programmed to generate successful abstracts or summaries, then the design of the software must be driven by a discourse theory of abstracts and abstracting. Put another way, a successful automatic abstracting machine can only be constructed once it is known why some abstracts are perceived to be more successful than others.

5

Review of the Library and Information Science Literature

5.1 Introduction

This chapter is organised in terms of three wh-questions (Quirk et al 1972), why, what and how, which are addressed by the Library and Information Science literature. Section 5.2 answers the question 'why write abstracts?' and describes the functions of abstracts and abstracting. Section 5.3 answers the question 'what are abstracts?' and discusses the different types which exist. Section 5.4 answers the question 'how does one write an abstract?' and reviews the normative standards and guidelines which inform prospective abstractors how to write abstracts. Some brief conclusions are provided in section 5.5.

5.2 Why Write Abstracts?

Abstracts are intended to save people time, and, indirectly, money. As information proliferates and becomes propagated ever more widely, it will inevitably be devalued unless it is kept in check. What is commonly referred to as the 'information explosion' means that one will never have sufficient time to read everything that one is interested in. Abstracts are a means of dealing with this problem by filtering information to more manageable size.

As well as conveying an informational skeleton of the source text, abstracts also help to overcome the language barrier and improve indexing efficiency (Borko and Bernier 1975: 7-9). Until machines can successfully translate unrestricted text between languages, translation will remain a very costly affair, but providing an abstract of a foreign language article in whatever language is expected to be the reader's mother tongue is a money-saving alternative. Turning to indexing now, clearly it would be undesirable for every

word of a text to appear in an index, but it would be equally unsatisfactory to index the words of the title only. Using terms from an abstract can often be a suitable intermediate step. Similarly, automated literature searches (or, looking for instances of a particular string, word, or phrase across millions of lines of text) are better performed on abstracts than on titles, which may not be particularly descriptive of their content, or on full text, which would be very expensive computationally. The importance of abstracts will become greater as more and more texts become more and more available to more and more people.

5.3 What are Abstracts?

All the references which have been consulted provide an explicit definition of the abstract. Five book-length accounts of abstracting and indexing have been used (Collison 1971, Maizell et al 1971, Borko and Bernier 1975, Cremmins 1982 and Cleveland and Cleveland 1983) together with some twenty or so articles. These cover editorial policies, attempt to formulate standards, collate various abstractors' instructions, or discuss miscellaneous aspects of the genre. On examining the various definitions of the abstract, a remarkable homogeneity presents itself. This is in contrast to the rules and recommendations the various authorities provide for the writing of abstracts, which are often contradictory.

While one authority states 'An abstract is defined as an abbreviated, accurate representation of a document which should be published with it and which is also useful in secondary publications and services' (ANSI 1977: 252), another makes the point that 'An abstract is an abbreviated representation of a document, without added interpretation or criticism. An abstractor must take a larger work, find its essence, and represent it concisely and accurately, without injecting the abstractor's personal biases' (ERIC 1980: 3).

More delicate definitions of abstracts are type-dependent and presuppose a richer understanding of the situation, needs, expectations, and goals of the readers of abstracts. These additional contextual factors will be discussed in the next section.

5.3.1 What Sorts of Abstract are there?

There are four major types of abstract: informative, indicative, critical and author-prepared. Informative and indicative abstracts are the most common.

- Informative Abstracts

Informative abstracts inform the reader of the findings of the original in such a way that readers will not necessarily need to refer to the source text: 'The informative abstract ... includes at least some of the more important data, facts, observations, or conclusions presented in the original source document. To some extent, and for some users, this type of abstract will provide useful information, making it unnecessary or less necessary to refer to the original document' (Speight 1977: 2). This makes the informative abstract a 'sovereign' text. Sovereignty 'denotes the self-supporting capacity of the text, its power to generate and develop a pattern of meaning, without reference to externals and without requiring of its readers any prior knowledge other than the common stock of experience' (Carter and Nash 1983: 130).

- Indicative Abstracts

Indicative abstracts, on the other hand, are less self-supporting and seek to indicate to the reader the basic contents without going into any details. They tell readers whether the material is going to be relevant for their purposes, and therefore whether or not to consult the original: 'An indicative abstract describes what a document "talks about" without recording "what it says." The indicative abstract does not include data, facts, observations, or conclusions that are discussed in the document. Instead it indicates that such data, facts, etc., are subjects discussed in the document' (Speight 1977: 2).

- Critical Abstracts

This type of abstract does not merely present the information, it also evaluates the information. Critical abstracts are generally frowned upon by the Information and Library Science communities and seem to be discussed mainly for reasons of completeness.

- **Author-Prepared Abstracts**

This type of abstract will be familiar to most academic readers - many will have prefaced their own journal articles with abstracts, or will have sent them to conference moderators. However, although common, author-written abstracts are often unhelpful (Collison 1971: 8 - 9), as sometimes the author's natural desire is to get the reader to read the whole document, so the abstract becomes more of an advertisement and less of a concise representation of the original.

The abstracts investigated in the present study are all assumed to be either informative or indicative. Examined below are the two major determinants of abstract type (ANSI 1977: 253, ASTIA 1959: 6 - 7, Weil et al 1963: 127, Borko and Bernier 1975: 13 f.f.): the needs of the readers and the type of the original.

5.3.1.1 The Needs of Readers

'Reader-slanting', or '-orienting', is 'simply the writing of an abstract of a given document in a manner best suited to the needs of a given audience ... A slanted abstract often highlights novel information in a document that is important only to the specific audience of the abstracter and is only incidental to the main subject of the document's author' (Weil et al 1963: 130). This is a difficult goal to achieve because, typically, different topics will be judged to be of differing levels of importance by different readers.

Another commentary states 'Abstractors write oriented abstracts when they emphasize some aspects of documents and deemphasize other aspects. Often this is done to meet the needs of a specialized audience - but not always. Some degree of orienting is natural, and indeed inevitable. After all, the abstractor is a product of years of education and of at least some experience in a specific discipline. His abstracts will almost inevitably reflect his background. A completely nonoriented abstract is too much to expect from anyone' (Maizell et al 1971: 93). In other words, an abstract can never be truly neutral, since the mere fact that the abstractor edits the source text is itself an act of evaluation.

The needs of readers also partly determine whether an informative or an indicative abstract should be furnished. Generalising from all the documents consulted, informatives are preferred if there is a discernible thread running through a text, while the use of indicatives is recommended for material of a

more disparate nature, and where attempts to provide an informative abstract may frustrate the reader.

5.3.1.2 Type of the Original

Choosing between an informative and indicative abstract also depends on a number of factors including the availability, length, and nature of the source text.

If an original is especially hard to obtain, it is unlikely that the readers of an indicative abstract would gain access to it and would therefore benefit more from an informative abstract.

In the case of a very long source text, on the other hand, it is considered preferable to provide an indicative abstract, which can show the scope of the original. This is considered better practice than reproducing an incomplete set of main findings.

The informative is thought to be the most useful option for those texts which conform to a rigid and stereotyped structural pattern, for example experimental studies (see Chapter 9 below which analyses two abstracts of a psychology article in terms of their generic structure). Where the generic structure of the source text is more disparate, or perhaps less well signalled, the indicative is preferred.

5.4 How does One Write an Abstract?

Relatively little space is devoted to the process of abstract writing in the normative literature. Collison 1971, for example, stretches to 120 pages, slightly under four of which provide any hint as to how one might set about writing a successful abstract. Similarly, Borko and Bernier 1975 is a 250 page work, but spends a similar amount of time discussing the body section of an abstract (pages 65 - 70).

The following discussion is divided into two. Section 5.4.1 sketches some of the rules provided by the various abstracting standards and guidelines. Section 5.4.2 then outlines what one authority considers to be the various steps in the abstracting process.

5.4.1 Rules

The rules discussed in this section include: how to write the 'topical sentence' (5.4.1.1); what structural elements need to be represented in an abstract (5.4.1.2); and other miscellaneous recommendations (5.4.1.3).

5.4.1.1 The Topical Sentence

The topical or 'lead' sentence is merely the first of the abstract, in which the abstractor is supposed to grab the reader's attention. In an article entitled 'Digesting for a multicompany management audience' which is concerned with chemical refining, J.C. Lane gives the following formula (1961: 62) for the successful topical sentence: 'Normally, the subject of the lead sentence is a tip-off to the subject of the whole item... [and] is in the form of a noun phrase having just enough words to identify the subject of the whole item. The predicate verb of the lead sentence almost always is an active, working one - telling what the subject does for or against the interests of refiners: e.g. saves (money), upgrades (octane number), causes (wear), etc. Normally, a noun, or noun phrase, or clause follows the verb, completing the statement about the subject and pointing out its significance'.

5.4.1.2 Structural Elements

Many of the authorities insist that there are a number of structural components which must be obligatorily represented in the abstract and should be readily identifiable. ANSI (1977: 253) states 'Readers in many disciplines have become accustomed to an abstract that states the purpose, methodology, results, and conclusions presented in the original document. Most documents describing experimental work can be analyzed according to these elements'.

Different sets of structural elements are investigated below in Chapter 9. Interestingly, the order in which structural elements are to be realised is claimed to be a function of the type of abstract to be written: Weil 1970, for example, maintains that the 'normal' ordering for an informative abstract can be transformed so that the 'results' section appears at the beginning of the abstract. Ordering is also investigated in Chapter 9 below.

Weil (1970: 353) also notes that certain 'collateral' information may be given: 'Include findings or information incidental to the main purpose of the document but of value outside its major subject area (e.g., modifications of methods, new compounds, newly determined physical constants, and newly discovered documents or data sources). Report these clearly, but in such a way that they do not distract attention from the main theme. Do not exaggerate their relative importance in the abstracted document', points echoed by ISO 214 (1976). However, neither authority makes clear where this additional information should feature.

5.4.1.3 Miscellaneous Recommendations

The ERIC manual (ERIC 1980) provides 'detailed rules' for the writing of abstracts. These include stipulations as to length, voice, tense, abbreviations and acronyms, and so on. It also provides guidance on higher-level phenomena: 'An abstract is always one paragraph long. The accepted rules about paragraph writing must be followed, especially those concerning coherence and unity. A coherent paragraph contains connected sentences, each following the other in logical order. An abstractor can avoid writing a paragraph that is nothing more than a series of sentences, each one summarizing a separate topic in the document, by intelligent use of transitional words and phrases' (ERIC 1980: 13). Unfortunately, the phrase 'intelligent use of transitional words and phrases' is neither explained nor exemplified.

The ERIC Processing Manual (1980: VI, 15) also discusses 'deletion rules':

- This report...
- The purpose of this report (or document) is ...
- It was suggested that
- It was found that
- It was reported that
- ... are discussed
- ... are given

ERIC argues that phrases such as these can safely be elided in the interests of economy; other authorities provide their own lists of 'redundancies' (see for example Kent and Lancour 1968: 24).

Interestingly, findings from Chapters 12 and 13 below challenge the wisdom of such deletion rules. Indeed there is evidence to suggest that readers, in the present study at least, respond favourably to these kinds of metadiscourse.

Weil et al 1963 maintain that correct spelling and grammar is to be used at all times: nominalisations are to be avoided as 'The noun form of verbs makes for dull reading. "Separating butadiene from butenes" reads better in an abstract than "the separation of butadiene from butenes"' (1963: 129); there must be no dangling participles; in matters of lexical choice the Anglo-Saxon form should take precedence over the Romance ("indexes", not "indices", 1963: 130); and the active voice is to be used whenever possible.

The suggested steps a writer should follow in the preparation of an abstract are outlined in the next section.

5.4.2 Procedures

The following procedure (Weil et al. 1963: 128) is typical of those provided by the various abstracting authorities: 'The following sequence has proved useful at Esso Research ... :

1. Read the introduction first if you are unfamiliar with the subject.
2. Read the author's abstract if one is included. Use it as far as possible if a further check of the document proves it to be a good one.
3. Review the summary and conclusions of the paper.
4. Scan the text for additional information, examining the captions of tables and figures.
5. Jot down marginal notes as you go along, and underline salient phrases and passages.
6. Then, and only then, write the topical sentence and the rest of the abstract (what was found and, where pertinent, why and how the work was done), plus the bibliographical citation'.

This provides some useful procedural context prior to the writing proper. However, stage 6, although it does provide some indication of the generic structure to be adopted (presumably 'Results - Problems - Methods' in this case), does not say how these elements are to be realised linguistically. This is typical of many of the published procedures: they are not sufficiently reified to be actionable.

5.5 Conclusions

The following conclusions are to be drawn from this chapter:

- The Library and Information Science literature provides some useful background information concerning the different types and functions of the abstract. A fundamental distinction is drawn between informative and indicative abstracts.
- However, there is little in the literature which can help answer the research questions listed in Chapter 1 above.
- The basic problem with the normative rules is that there is no guarantee that following them will result in abstracts which will be perceived to be successful. This is often because the types of process which are most easily verbalised (for example, punctuation, typography, spelling and so on) are very often not particularly important determinants of perceived quality. Chapter 8 below reviews the different reasons judges believe underpin their preferences; the most frequently mentioned are higher-level phenomena, such as 'coherence' or 'flow of discourse'. These are criticisms which are altogether harder to address.
- The procedures documented in the literature do not actually inform abstractors *how* to write a successful abstract; they merely instruct them to do so. A typical instruction is: 'write concisely, write clearly and understandably'. However, this is a less than useful piece of advice because it refers to concepts which are impressionistic and not well understood linguistically. In other words, the instructions are not sufficiently well spelt out to be directly actionable.
- No attempts have been made to assess how successful different abstracts are perceived to be by genuine consumers. There have been one or two fact-finding exercises, notably Borko and Chatman's 1963 survey, but these are merely ethnographies of the authorities themselves, and have little to do with the perceptions and expectations of readers. In effect, the data outlined in this chapter is not well 'grounded' (see Chapter 2 above) in the sense that it is not based on naturalistic observations of real-life abstracting practice.

Part Three

Data Collection

Introduction

The following brief introduction has two purposes: firstly, it assesses the degree to which the literature reviewed in Part Two above can be used to answer the basic research questions stated in Section 1.5 above; and secondly, it provides an overview of the data collection phase of the present study which is detailed in Chapters 6, 7 and 8.

All of the research questions this thesis is intended to answer concern the notion of success as perceived by the reader. However, in the literature, readers' perceptions and expectations of the quality of abstract writing are seldom sought out.

This is certainly true of the work in Artificial Intelligence where a great deal of effort is expended constructing systems but little is spent evaluating them. In fact, one of the conclusions from Chapter 4 above was that, rather than merely being used to evaluate such software, insights from discourse theory should be used to design the systems from their inception, the argument being that one cannot hope to build a successful automatic abstracting system without first understanding what constitutes a successful abstract.

The normative literature from Library and Information Science similarly makes little or no appeal to the intuitions of genuine producers and consumers of abstracts. This makes the advice they have to offer of only limited utility. As mentioned above in Chapter 5, their advice mainly concerns those types of process which are most easily verbalised (for example, punctuation, typography, spelling and so on); the things which seem to matter most to real users of abstracts are at a much higher level and are consequently much less well understood.

The psychological literature is more useful in that naturally occurring texts are made the subject of enquiry, but these are rarely investigated using precise

linguistic analysis. Although there are considerably fewer studies of abstracts and abstracting in Linguistics than there are in Psychology, these are of more direct relevance to the aims and objectives of this research. Drury 1989 (see Chapter 3 above) was particularly instructive in that she used a Systemic-Functional framework to differentiate between three summaries of differing levels of perceived success. However, her success judgements were not directly related to the summaries, so caution has to be exercised when trying to reconcile her internal measures with her external measures.

Because the secondary data reviewed in Part Two proved not to be sufficiently rich to answer the basic research questions of the present study, Part Three describes the primary data collection which had to be carried out. Two types of data had to be collected: a set of comparable, naturally occurring abstracts and a set of judgements and intuitions concerning those abstracts.

Chapter 2 above provides a discussion of the general research design of the present study in which fieldwork (section 2.3.1) and surveys (2.3.2) were considered to be the best means of collecting such data.

Fieldwork was selected as the most appropriate method to collect the texts, since the principal requirements for the abstracts were that they had to be natural, grounded, and had to be linguistically reconcilable with the success judgements. Chapter 6 below shows how these requirements were satisfied by the field study at Brighton Polytechnic.

Questionnaires were selected as the most appropriate method to collect the subjective data. Ideally, the subjective data should also have been naturally occurring, but there was not sufficient available solely from the field (for example, in their evaluation of the students' abstracts, the lecturers rarely discuss particular instances of writing). Consequently, Chapter 7 describes how four different kinds of subjective data were collected: intuitions concerning the abstracts viewed as products; intuitions concerning the source material viewed as products; intuitions concerning the processes involved in writing the abstracts; and intuitions concerning the processes involved in reading the originals and selecting the information to be represented in the abstracts.

Finally, Chapter 8 entitled 'Results of the Data Collection', assesses how successful the primary data collection was and provides a catalogue and commentary of the various data collected.

6

Fieldwork Data

6.1 Introduction

This chapter describes the fieldwork phase of the study. Firstly, the objectives of the fieldwork are stated (6.2). Section 6.3 then discusses how the fieldwork was planned so as to meet those objectives: how the choice of site was made (6.3.1); how the fieldworker's presented role was decided upon (6.3.2); and how long (6.3.3) and how focused (6.3.4) the study would be. The actual fieldwork on which this study is based is described in Section 6.4 onwards. Details are given both of the abstract writers (6.4.1) and of the source material they were asked to summarise (6.4.2). Conclusions are given in 6.5.

6.2 Objectives of the Fieldwork

The most basic objective of the fieldwork was to collect a number of abstracts for later analysis. However, to 'facilitate realistic theoretical interpretation' (Brewer and Hunter 1989: 160), those abstracts had to be both natural and grounded (see Chapter 2). To be fully natural, the abstracts had to be written by real people for real readers, and had to be surrogates for real source texts. For the abstracts to be grounded presupposed 'a commitment to get close to the program in its natural setting, to be factual and descriptive in reporting what is observed, and to find out the points of view of participants in the program observed' (Patton 1987: 74). Collecting grounded abstracts, then, demands not only the accumulation of abstracts, but also records of the context in which they were embedded; Part Three presents, therefore, a catalogue of the collected abstracts, a certain amount of field details, various participants' views of the abstracting activity, and, for reasons that will be discussed in the next chapter, further judgements of the quality of the abstracts made by people who were *not* part of the field described in this chapter.

As well as having natural and grounded qualities, the abstracts had to be in sufficient quantity to enable comparisons to be made between examples judged to be of different degrees of success. For the comparisons to be fair, certain constraints had to be satisfied. Firstly, there had to be a number of *different* abstracts written for the *same* source text; this ensured that the different perceived levels of success would be due to differences in the quality of the writing, and would not be affected by differences in the subject matter of the source texts. Since it would be extremely unlikely for any one abstractor to write more than one abstract for the same source material, it was considered necessary to use multiple subjects. Secondly, different types of source material had to be used to avoid the danger of formulating hypotheses which accounted for one narrowly defined text-type only. This was thought all the more important since it is widely reported (see Chapter 5 above) that different types of subject matter require different modes of abstracting.

To sum up, then, the fundamental requirement of the fieldwork was that texts had to be collected which were as grounded and natural as possible. However, it was anticipated that a certain amount of control might have to be introduced to enable fair and precise theory building. This will be reviewed in the next section.

6.3 Planning the Fieldwork

The fact that fieldwork is widely reported to be a largely unpredictable, inductive and learning experience should not be taken to mean however that planning is not an essential prerequisite of a successful field study. This section therefore discusses decisions that were taken prior to the data collection. These concern: the site itself (6.3.1); the presence of the researcher (6.3.2); the duration of the study (6.3.3); and its level of focus (6.3.4).

6.3.1 Planning the Ideal Site for the Fieldwork

Marshall and Rossman (1989: 54) list four criteria for site selection. For them, the ideal site is one in which:

- entry is possible;
- appropriate data can be collected;
- the researcher can devise an appropriate role (see next section); and
- collected data is credible or trustworthy.

In practice the ideal is seldom attainable, and a certain amount of compromise must be allowed.

For the site to meet all of these criteria meant that fieldwork had to be conducted in an educational establishment (the choice of which particular establishment is discussed 6.4 below). The reasons for preferring the educational domain will now be discussed in relation to the four criteria above.

Firstly, it was assumed that people in a teaching environment would be more sympathetic to the educational aims of the research, more used to collaborating with external researchers, and therefore *more likely to allow entry* into their province.

Secondly, the data available in an educational setting was thought particularly well suited to the aims of the research for the following four reasons:

i) the data would be likely to exhibit *varied levels of perceived quality*. This would mean that abstracts judged more successful could be compared with abstracts judged less successful. This is particularly advantageous since the assumption underlying this research is that the best way of characterising a successful abstract is by showing how it differs from an unsuccessful abstract;

ii) to be strictly comparable, however, ideally one would require that abstracts differed only in terms of their perceived success. In other words, one would want to ensure that different levels of perceived quality were due to differences in the writing and nothing else. It would be not be fair, for example, comparing an abstract written by an experienced paid abstractor from an article on quantum physics with an abstract written as an eighth grade class exercise from an account of the American Declaration of Independence. The educational setting afforded an *opportunity to standardise* both on the experience of the writers, and on the source material to be abstracted, since numbers of students are often asked to carry out similar, if not identical, tasks;

iii) further, such work is *typically evaluated by teachers as a matter of course*. This accords with advice given by Herman, Morris and Fitz-Gibbon (1987: 34),

who say 'try to find useful information which is going to be collected anyhow. Find out which assessments are given as part of the program or routinely in the setting';

iv) in spite of the fact that novice abstractors can hardly be viewed as being experts, their experiences of how they fared with the new task are nevertheless worth scrutinising, since the *identification of problems encountered* is a useful stepping stone to the formulation of useful solutions.

Thirdly, students are well used to having their work evaluated, and so it was thought that the presence of an additional 'assessor' would lead to *fewer observer effects* than would have been the case in most other situations. This is not to say that the researcher's presence had no effect, however; issues of observer presence will be returned to in the next section.

Fourthly, it is acknowledged that poor sampling decisions can threaten the credibility or general trustworthiness of the data (Brewer and Hunter 1989: 161), and so the field had to be chosen to be *representative of wider abstracting practice*. Therefore, rather than trying to gain access to a professional abstracting company, which may well have had its own peculiar house style, or subject specialisms, the educational sphere was considered more appropriate, since one would expect the training provided to be *general enough* to equip students to apply for different types of abstracting jobs.

Educational establishments offering courses in abstracting were therefore targetted as being the most appropriate type of site for the fieldwork. The means of narrowing down the search to the particular case study used for this research is described in section 6.4 below.

Fieldwork planning does not end once the choice of site has been made, however. Patton (1987: 74 - 82) lists five continua along which fieldwork may vary, the first three of which pertain to the received image of the fieldworker (6.3.2), the fourth to the length of the study (6.3.3), the fifth to its level of focus (6.3.4). The following three sections discuss the further choices pertaining to each of these dimensions.

6.3.2 Planning an Appropriate Role for the Fieldworker

The first of the three continua which pertain to the role of the fieldworker

(Patton 1987: 74 - 76) concerns the degree to which the researcher takes part in the everyday activities of the field. Thus, at one end of the continuum the fieldworker *participates* in fieldwork life, at the other the fieldworker merely *observes*. Both extremes have advantages and disadvantages. Being an observer allows the fieldworker the most time to concentrate on monitoring fieldwork life. Being a participant, on the other hand, enables the fieldworker to become an insider, to feel as well as see events in the field. In the abstracting context, participating would have entailed the researcher writing his own abstracts and having them evaluated by the Polytechnic staff.

As it was assumed that the researcher would later be analysing the students' abstracts and relating these to success judgements, it was decided in the interests of objectivity not to participate directly. This meant that the researcher did not read any of the source material, so that his understanding of the original texts should not bias his attitude to the abstracts. It also meant that the researcher could better empathise with normal users of abstracts, who do not read the originals in the normal course of events. Another significant advantage of maintaining the role of observer is that fieldworkers can formulate contingency plans if need be, and are generally more free to take in what is going on around them: to take notes concerning the events or surroundings; to record the order in which the students finished their work; to find replacements for unexpectedly defective microphones, and so on. It was therefore decided to maintain the role of observer throughout the fieldwork study.

The second of the three continua (Patton 1987: 76 - 77) pertains to the way in which fieldworkers portray themselves to others. At one end of the continuum, fieldworkers can conduct their observation in a fully overt manner, at the other in a fully covert manner. The main advantage of being overt is that, pending the third decision (see below), one can ask very specific questions concerning the study. The main advantage of remaining covert is that one minimises bias in the form of observer effects. There are also very large questions of ethics involved. It was decided mainly for ethical reasons to observe proceedings in plain sight, as it were. Although it must be assumed that data collected by means of overt observation will inevitably contain observer effects, these can be controlled to some extent by careful consideration of the next planning decision.

The last of the three continua (Patton 1987: 77 - 78) pertains to the way in which fieldworkers portray the purpose of their fieldwork to others. At one end of the continuum lies full disclosure, at the other complete non-disclosure.

Alternatively, there may be contexts in which it is appropriate to provide false information. For the purposes of this study, it was decided to follow the compromise position often adopted in linguistics, in which subjects are not told any details of the study until after the fieldwork has been completed. As it turned out, therefore, both the lecturers and the students were 'kept in the dark' until well into the second term. This meant that observer effects, though present, were controlled to a certain extent.

The next section discusses the fourth of Patton's five dimensions along which fieldwork can vary, that according to length of study (Patton 1987: 79).

6.3.3 Planning the Duration of the Fieldwork

Following on from what was said in the last section concerning observer effects, one of the problems with a one-shot overt observer case study is that subjects do not have time to get used to the observer, and to the idea of being observed. It was thought preferable therefore that the same group of students should be observed on a number of different occasions, even if only separated by a short amount of time.

Sometimes another way of combating the problem of observer bias can be to introduce the observer as early as possible, certainly before the observation proper begins. Section 6.4 below discusses how subject 'acclimatisation' was attempted.

6.3.4 Planning the Focus of the Fieldwork

Planning the focus of the fieldwork (Patton 1987: 80) refers to which particular aspects of the fieldwork are to be concentrated upon. This is less relevant in the present study, however, since one of its major motivations is to discover which are the most important variables. In cases where there is no well developed theory, or where the context is not well understood beforehand, it is appropriate to maintain a broad focus, or to 'keep an open mind', so as not to disallow anything which might later prove interesting in the analysis phase of the work. It was therefore decided not to pre-judge which would be the most interesting areas for investigation, but to proceed inductively, narrowing down only once patterns in the data became evident.

The researcher tried to maintain a broad focus by remaining 'open' to the

data in two ways: firstly, by controlling participants as little as possible; and secondly, by collecting as much field data as was practical, rather than concentrating on any one aspect.

6.4 The Brighton Polytechnic Case Study

This section describes what actually happened in the fieldwork phase of the project, so that comparisons between the ideal, discussed in the last section, and the actual can be made.

The choice of educational establishment had to be made in as principled a way as possible, and so the selection was made firstly, by identifying those which offered formal teaching in abstracting, and secondly, by gauging how likely the staff were to help. Relevant addresses were obtained by using an online database called "PICKUP", a directory of courses available at polytechnics, universities and colleges. Searching for the keyword "abstract" in the Information and Library Science section of the database resulted in fifteen relevant sites being flagged. Letters were despatched to all of these, and, of the ten replies, three contained invitations to visit. These three potential sites were evaluated according to the following criteria: the nature and scope of the training given in abstracting; the number of students undergoing this training each year; and finally, the perceived willingness to help of the members of staff.

Brighton Polytechnic was finally selected as being the most appropriate site for the fieldwork. The Polytechnic runs a three year full-time B.A. Honours degree in Library and Information Studies. At the time of the data collection, two members of staff were responsible for abstracting: the first had taught abstracting to many successive years of students; the second had recently joined the Department, having been employed in previous jobs as a professional abstractor herself. These two members of staff were thought to be particularly good informants.

The fieldwork was conducted over a period of four or five weeks in the winter of 1988. This meant that the disadvantages of the one shot study were avoided. The different fieldwork sessions are timetabled in Appendix 6.1: the first an introduction to abstracting, and to the class exercises; the second a practical writing session in which students were asked to write their own abstracts for selected texts; the third an evaluation of their work in the previous session; the fourth more practical work, this time from material covering a different

subject; and the fifth, an evaluation of the previous session, together with some concluding comments. Each of these sessions was tape-recorded (see Appendix 6.2) as it was thought that this might provide valuable supplementary data of an evaluative nature concerning both the teachers' appraisals of the various abstracts, and the students' reactions to the task (these data are reported on in Chapter 8).

The observer attended lectures and evaluation sessions prior to and during the more formal abstracting exercise in order to make himself slightly more familiar to the students. As mentioned above, the observer did not discuss the purpose of the fieldwork until well after the study when a lecture was given to the lecturers and students detailing the aims of the work, the anticipated means of analysis, and some techniques for automated information retrieval and management.

The remaining sections in this chapter give details of how the abstract writers were selected (6.4.1), how the source texts were selected (6.4.2), and finally, in the concluding section, a catalogue is provided of the abstracts themselves. These represent the textual data on which this study is based.

6.4.1 The Abstract Writers

Second year students were chosen to be the subjects of the study, as abstracting was taught in the normal course of events as part of the second year of the degree course. Another advantage of observing that particular year was that it was entirely coursework-based and the absence of exams usually makes for a more relaxed year. In spite of this, because of the experimental nature of this research, it was agreed between the lecturers and the researcher at the outset that the students' abstracting performance would not form an official component of their coursework.

As section 6.4.2 below assumes that the subjects can be thought of as being subject specialists in library and information work, a brief overview of their first year syllabus will be given, together with some brief biographical details. The second year of the degree course 1987 - 1990 consisted of 34 students, 29 female and 5 male. Most had come straight from school, but there were a small number of mature students. At that time two GCE A levels were required for entry, with a pass at GCSE / GCE O level English Language. All students take five modules: information environment, management, information science, information communication, and introduction to

computers; the last two are assessed entirely by coursework, the others by a 3:2 ratio of exams to coursework. Students then sit three three-hour exams in the third term, and proceed to a one-month full-time placement at an information unit or library most usually somewhere within the U.K.

The data observed in the fieldwork constituted part of the third module, Information Science; the modules will not be detailed here, suffice to say that they provide broad background knowledge and, as such, form part of what is assumed to be necessary for abstracting material from information and library science. As will be seen in the next section, half the total number of source texts used concern issues or theories from this subject area.

6.4.2 The Source Texts

Above it was stated that some means of control had to be introduced into the otherwise natural proceedings so that fair and useful comparisons could be made. In this section, therefore, clear differentiation will be made between what was decided by the lecturers, and what was requested of the lecturers by the researcher.

To meet the objectives of the fieldwork, there had to be scope for inter-text comparability, and observer effects had to be controlled as far as possible. Before the way in which the researcher influenced the events in order to meet these criteria is described, some details will be given of what would have happened in the normal course of events.

The lecturers thought it a good idea to use two different types of source material; this meant that each student would be expected to write two abstracts. Both of the source types had to be capable of being summarised by the students, subject familiarity being a prerequisite for successful abstracting. Consequently literature from the Information and Library Science fields were used. These texts were supplemented by General Knowledge material, texts which the students would have no difficulty comprehending. One of the lecturers further stated that it was not only the subject matter of the original which affected the writing of an abstract, but also its length and type. They therefore decided to have two types of each class of source material: General Knowledge originals in the form of popular magazine articles and in the form of newspaper items; and Information and Library Science originals in the form of learned journals and in the form of book chapters.

This decision making was done in front of the researcher, but was not in

any way influenced by him. The researcher did, however, make two explicit requests to further meet the objectives stated above.

Firstly, the researcher asked to have two different instances of each sub-type of original, so, for example, instead of having one Information Science journal article being abstracted by a certain number of students, two would be abstracted, each by half that certain number. This was done to allow comparability between different instances of the same genre.

Secondly, the researcher asked for the Information and Library Science material to be abstracted in the second practical; this was done because the subject specialist material was felt to be the more interesting data, since pilot data collection from the commercial sphere had revealed that the vast majority of abstracts are written for highly specialist original material. Placing this data second meant that it should have been less prone to observer effects.

All the texts were chosen by the lecturers. In fact the researcher did not have sight of these until they were given to the students at the beginning of the practical sessions. The eight different source texts used are referenced in Appendix 6.3. All were quite recent; in the case of the newspapers, the articles were taken from the issue one day prior to the experimental session. The other two general interest items came from popular periodicals with relatively high circulation and were taken from the previous month's edition. Two of the specialist texts were from specialist library and information science periodicals, one from 1984, and other from 1988. The remaining two specialist items were book chapters taken from Vickery and Vickery 1988.

6.5 Conclusions

Given that fieldwork is a notoriously precarious enterprise, this phase of the project was particularly successful. Many different types of data were collected: the data comprising opinions and judgements will be discussed in the following two chapters in Part Three.

The textual data collected was rich, varied and, above all, allowed abstracts judged to be successful to be compared with abstracts judged to be less than successful. Further, a number of different source texts were used to see to what extent the type of original influenced abstracting practice.

In total 42 abstracts were written, out of an absolute maximum of 68 (34 second year students all writing two abstracts each). 13 students wrote two abstracts, 16 students wrote one, and 5 wrote no abstracts. The distribution of

source texts to abstracts is shown in Figure 6.1 below:

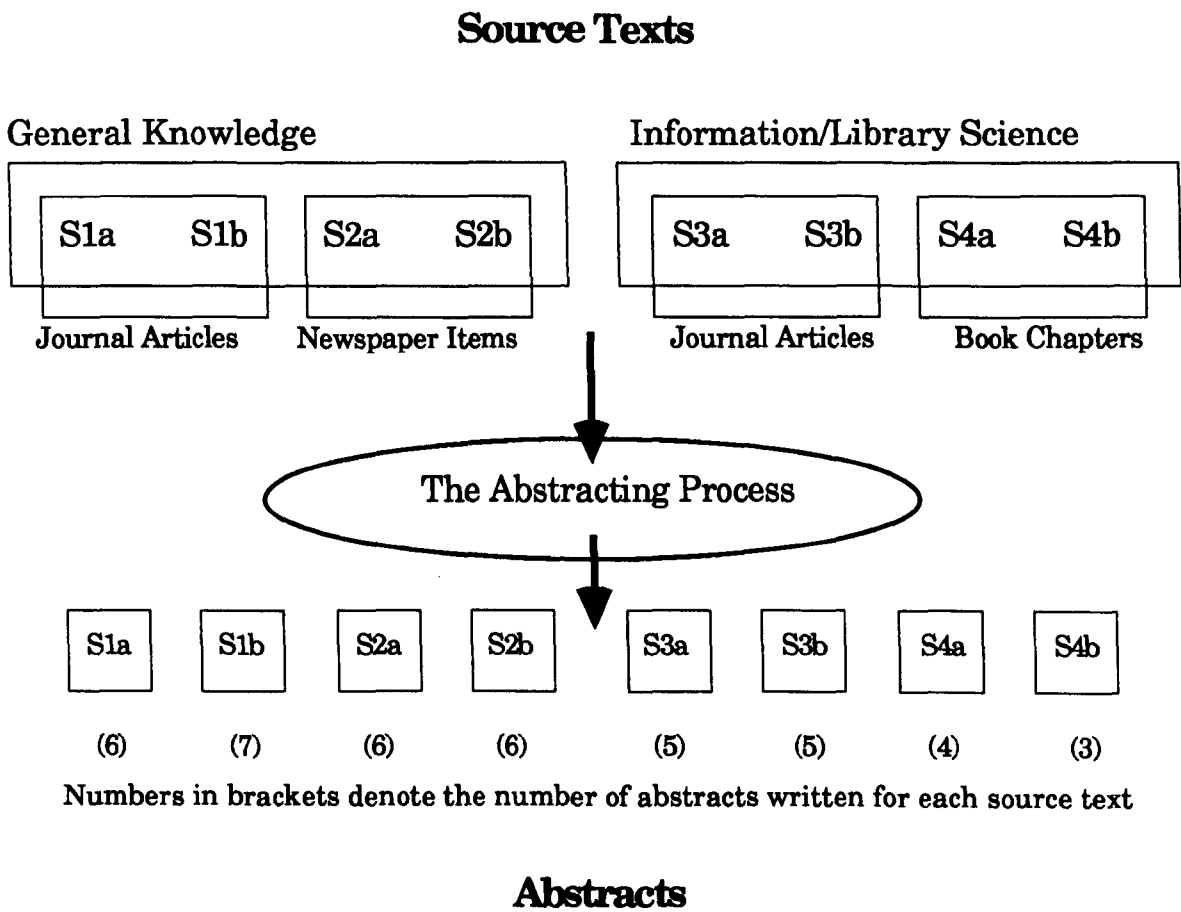


Figure 6.1: Textual Data Collected from the Field

In Figure 6.1, both source text and abstract are denoted by an S- reference devised by the lecturers: there are four different types of source text (1 - 4), with two different instances of each (a and b). The number of abstracts written for each source text is shown in brackets and varies from three (S4b) to seven (S1b).

To give an indication of the amount of textual data collected, the number of words contained in each of the abstracts is shown in Figure 6.2 overleaf.

Total Number of Words in each General Knowledge Abstract

Abstract Version: A	B	C	D	E	F	G	
S1a Erosion:	153	207	354	147	130	292	-
S1b Cocaine:	145	125	102	97	83	331	277
S2a Brain Drain:	110	97	102	166	94	170	-
S2b Nature/ Nurture:	130	199	178	239	135	169	-

Total Number of Words in each Information/Library Science Abstract

Abstract Version:	A	B	C	D	E
S3a Distinct Personality Type:	651	323	230	191	127
S3b Tanzania:	198	103	237	274	277
S4a Vickery and Vickery, Chapter 1:	268	272	238	471	-
S4b Vickery and Vickery, Chapter 2:	199	265	111	-	-

Figure 6.2: Table of Abstract Lengths

The next chapter, Chapter 7, shows why additional data in the form of judgements had to be collected, and reports on how and why evaluative opinions were sought from both producers and consumers of the abstracts.

7

Survey Data

7.1 Introduction

This chapter describes how data in the form of opinions and judgements were collected to complement the textual data discussed in Chapter 6 above. As Chapter 8 will show, the teachers' evaluations of their students' abstracts which were collected as part of the fieldwork are very general, and rarely discuss particular instances of writing. Further, because the abstracts were not given a formal marking, there was no reliable means of discovering how and why some were preferred over others. Consequently judgements had to be collected in a more controlled way using sets of judges, each asked i) to perform a very specific ranking procedure, and ii) to provide reasons for their preferences. In addition, the chapter also discusses how other more general opinions were sought, not merely from model abstract consumers, but also from the abstract writers themselves, and from their teachers.

Section 7.2 discusses what the requirements were for the design of the survey, and reviews the four different kinds of data which had to be collected (7.2.1 - 7.2.4). Section 7.3 concerns the data collection proper: section 7.3.1 provides the rationale underlying the choice of informants; section 7.3.2 outlines the questionnaires one by one, and argues what each question was supposed to achieve; issues of graphical layout are treated in section 7.3.3, together with a number of remarks concerning the administration of the survey; finally, section 7.3.4 briefly reviews the way in which the questionnaires were pilot tested.

The six different questionnaires used in this part of the research - QJ1, QJ2(1), QJ2(2), QJ3: Slip, QJ3: General, and QJ4, 'Q' standing for 'Questionnaire' and 'J' for 'Judgements' - are presented in the Appendices (see Appendices 7.1 - 7.6).

7.2 Survey Data Requirements

This chapter discusses the four types of survey data which had to be collected for this stage of the study. Unlike Chapter 6 above, in which the data collected was natural and grounded, the data discussed in this chapter came not from the field, but was *elicited* to compensate for the lack of indigenous naturally occurring, reliable and systematic success judgements.

This section, which is an introduction to the following four sub-sections (7.2.1 - 7.2.4), discusses why four different kinds of elicited data were collected. The figure four derives from the desire to consider: abstracts viewed as products; the source material viewed as products; the processes involved in writing the abstracts; and the processes involved in reading the originals and selecting the information to be represented in the abstracts. Before these four types of data are treated individually, however, it is appropriate that the product/process distinction be discussed in a little more detail.

Over the last few years there has been a hotly disputed debate in textlinguistics between those researchers who maintain the discourse-as-process position, and those who prefer that of text-as-product (the two stances are conveniently summarised by Brown and Yule, 1983: 23 - 25). Written text is most usually presented as a *fait accompli*, a finished, tidy product, something reified that can be handled and scrutinised. Even when discourse analysts turn their attention towards spoken language, they typically make use of a transcription, itself a finished product, albeit somewhat less tidy. The preoccupation with textual artefacts has led many linguists into carrying out their analyses with little or no regard for those processes which enable language users to produce and comprehend utterances on particular occasions of use; the argument concerns whether or not this constitutes legitimate practice.

The multimethod approach adhered to in this thesis (see Chapter 2 above) demands that both product and process concerns be addressed, that both abstracts and abstracting be investigated, in other words. One good way to address product concerns is by eliciting success judgements from informants: this is because in so doing informants have to read the abstract and are therefore attending to it primarily in its guise as a written product (see Halliday 1985: xxii - xxiii on the differences between system and text, speaking and writing, and process and product). One good way to address process concerns, on the other hand, is by considering sets of 'how' questions primarily concerning the writing, rather than the reading, of abstracts. This

is because, as Halliday argues, issues of process are *more apparent* when, as ordinary language users, we construct our own written texts, as opposed to when we read other people's. Process questions to do with abstracting might therefore include:

- how does the abstract writer know how to select information from the source material for inclusion in the abstract?;
- how does the abstract writer know how to order this information?;
- how does the abstract writer know how to put this information into words?.

Unfortunately it is notoriously difficult to answer questions such as these, precisely because it is hard to observe the functioning of such processes on normal occasions of use.

Psycholinguists investigate process issues typically by using controlled experiments in which subjects are closely monitored while they carry out some simple task set by the experimenter. However, the use of experiments goes against the spirit of naturalism, in which artificiality is to be avoided.

There is, then, the following dilemma: this thesis borrows much from the naturalistic style of enquiry, which urges researchers to interfere as little as possible with the everyday goings-on in the field; however, this thesis is also highly committed to the multimethod approach, which urges that both product and process issues be addressed. Arguably, these two positions are incommensurable, since it is difficult to see how process can be investigated independently from product in anything but the most controlled, and therefore non-natural, of circumstances.

There is a partial solution to this dilemma, which is as follows. The view taken in this thesis is that process insights can indeed be had from the analysis of naturally occurring products, an opinion shared by a number of researchers: see work such as Gorrell 1983, Odell 1983, and Witte and Cherry 1986. But, to be able to combine the two, informants' intuitions concerning process-type activities had to be collected in a way which would not affect the natural course of the abstracting activity. In the present study this meant asking process as well as product questions, *but only once the formal class exercises had been completed*.

As stated above, questionnaires were designed to elicit four types of opinion, the four being the result of the two simultaneous choices, one between product

and process, the other between abstract and source text. These will be discussed in the next four sub-sections (7.2.1 - 7.2.4 below).

7.2.1 The Requirement to Collect Source Material Process Data

One requirement for this phase of the data collection was to assess *how* an original's 'abstractability' affects the process of writing an abstract. However, in order not to introduce too much of an element of control into the abstract writing activity, abstract writers were asked this question by means of questionnaire once the actual exercise had been completed.

As section 2.3.2 in Chapter 2 above pointed out, questionnaires may incorporate open or fixed-choice questions, or both. How-type questions are better realised as open questions, since at this point in the investigation it is the identification of variables which is important, rather than the attempt to quantify their effects. However, researchers must be aware that it is possible to ask questions which are 'too open'; answers to these may lack focus, and can be so vague that little progress is made. In linguistics this problem is compounded by the inordinate amount of time it takes to analyse even short stretches of language.

For these reasons it is sometimes prudent to elicit informants' opinions on particular sections of texts, rather than on texts in their entirety. This anticipates the analysis phase when time is at a premium by focusing the attention on more manageable amounts of data. This need not necessarily mean that informants are inveigled into having opinions they would otherwise not have had, since the informants themselves can be requested to make the appropriate selection.

This technique can be especially powerful when used to select those parts of a text which are considered by informants to exhibit both high and low degrees of whatever it is that the investigator is interested in. So, for example, asking informants to point out those sections of the original they found hardest and easiest to summarise provides a neat means of focusing the analysis on units it may be useful to compare and contrast. Another useful strategy used throughout all the questionnaires was to ask informants to back up their opinions by providing reasons for their evaluations.

Section 7.3.2 below describes how source material process data was collected by means of three questionnaires: 'QJ2(1)', 'QJ2(2)' and 'QJ3 General' (see

Appendices 7.2, 7.3 and 7.5).

7.2.2 The Requirement to Collect Source Material Product Data

As is widely reported in the literature on abstracting, the nature of the original material affects how it is to be abstracted (Chapter 5 above). Further, as has been pointed out in both computational and psychological research in studies of summarising (Chapters 3 and 4), it is not possible to construct a summary for a text which is wholly incoherent. It is to be expected, then, that the success of an original more generally will have a strong bearing on the ease with which it may be abstracted.

Another requirement for this phase of the data collection was therefore to assess the 'abstractability' of the original material, in order that fair comparisons could be made between abstracts which summarised different source texts. Also, it was felt that knowledge of this kind could help explain why some material presented particular difficulties for abstractors.

For reasons that will be described below in section 7.3.2, source material product data was collected by means of two questionnaires: 'QJ3 Slips' and 'QJ3 General' (see Appendices 7.4 and 7.5).

7.2.3 The Requirement to Collect Abstracts Process Data

Questions relating to the abstracting process concern those strategies which abstractors employ in the course of their writing. As has already been discussed, an early decision was made to investigate these solely through the perceptions of the writers themselves.

The fundamental requirement of this stage of the data collection was therefore to collect useful and interesting abstracting process insights. This meant that a careful balance had to be struck between wholly open and wholly structured questions. Given that such processes are to a large extent 'hidden', relying on wholly open questions (for example, 'How did you write your abstract?') runs the risk of obtaining unhelpful information of the 'I don't know; I just did it' type. But relying on wholly structured questions runs the risk of prejudicing the issue, or, put another way, of having to guess at the

answer in order to ask the question. In such situations a compromise position can be taken by guiding informants, as it were, without necessarily having to lead them all the way.

The most appropriate technique therefore for eliciting the required data was to ask questions about specific strategies, while still allowing informants freedom in the way they answer. However, narrowing down the scope of enquiries must always be done in as principled a way as possible; here Hasan's pioneering work on generic structure (1985/1989a) provides a principled means of focusing on specific abstracting strategies by motivating a tripartite examination of obligatoriness, order and realisation. First of all, it is reasonable to assume that for an abstract to be (perceived to be) successful, there will be a certain set of informational elements which must be obligatorily present. These elements may simply be a selection of those represented in the original, or may generalise on points made therein, or may paraphrase certain facts. Further it seems reasonable to assume that for an abstract to be (perceived to be) successful, these selected elements should occur in a certain order. It would be useful therefore to know how this order is determined; do the selected elements feature in the same relative order as they do in the original for example? The third assumption is that the selected elements should be realised, or put into words, in a certain way, so it would be useful to discover how writers decide between different candidate modes of expression.

Section 7.3 below shows how this tripartite design was exploited in the design of the three questionnaires which were used in this phase of the data collection, 'QJ2(1)', 'QJ2(2)' and 'QJ3 General' (see Appendices 7.2, 7.3 and 7.5).

7.2.4 The Requirement to Collect Abstracts Product Data

The last and most important type of data to be collected was evaluations of the successfulness of the abstracts made by people who regularly read and make use of abstracting facilities (see Chapters 1 and 2 above). This is the most important type of data because it most squarely addresses the aims and objectives of this thesis, the emphasis being on 'discourse theory', as defined above in Chapter 1. Informants' evaluations concerning the perceived quality of the abstracts viewed as products (i.e. 'external measures', in the terminology of Chapter 1) can be used to generate hypotheses to help discover to what extent certain linguistic factors (i.e. 'internal measures', in the

terminology of Chapter 1) influence the perceived success of an abstract.

The requirements for the success judgements data were two-fold: firstly, there had to be a quantitative means of measuring success, to facilitate rigorous hypothesis testing - informants were therefore asked to rank various abstracts according to how helpful they believed them to be; secondly, there had to be a qualitative follow-up, in which informants were invited to provide reasons for their preferences.

The aim was to investigate to what extent the qualitative reasons given by the informants explained their quantitative orderings, and, further, to decide which additional linguistic hypotheses had to be posited in order to distinguish good abstracts from bad.

As has been pointed out above in Chapter 2, this part of the data collection was primarily intended to generate hypotheses. At this stage it would not have been proper to ask more structured types of question such as 'was the language clear?', or 'did the ideas flow together in satisfactory fashion?', because at that time it could not have been guaranteed that such questions had any direct bearing on success. They were not 'grounded' questions, in other words.

In section 7.3 below, the exact means of collecting the abstracts product data is discussed. In total, four questionnaires were used: 'QJ1' and 'QJ4' (see Appendices 7.1 and 7.6) were administered to models of abstract consumers, and 'QJ2(1)' and 'QJ2(2)' (see Appendices 7.2 and 7.3) were administered to the abstract writers.

7.3 Collecting the Data: Questionnaire Design and Administration

Having set out the different types of opinion which were required for the study, this section discusses exactly how these were collected. The next four sub-sections discuss fundamental issues in the design and administration of the questionnaires: deciding who to use as informants (7.3.1); deciding which questions to ask, and how to ask them (7.3.2); deciding how to select and present the texts, deciding what the informants were to be told, and deciding how best to present the questions graphically (7.3.3); and, the use of a brief pilot (7.3.4).

7.3.1 Selecting the Most Appropriate Informants to Question

Triangulating between informants is potentially beneficial, but care must be taken not to overburden any one set of judges with work. Although in practice the constraint to use informants as economically as possible means that some questions can be put to different kinds of informant, others can be put to one kind only. This section shows how informants were selected to provide the required data, each type being particularly suited to answering certain kinds of question.

As for the source process data (section 7.2.1 above), the students were asked to identify those sections of the originals they found hardest/easiest to summarise, as discussed above. However, because the student producers cannot be considered expert abstractors, this is one instance where informant triangulation was thought to be most definitely worthwhile. Their intuitions had to be checked against those of more experienced abstractors. The easiest way of doing this was to ask their teachers, as they can be considered professional abstractors themselves. They had not actually written any abstracts for the originals, however, so their intuitions of how they would have gone about the task were collected instead. This data is speculative in nature, so an attempt was made to compensate for this by collecting it in two different ways: questions were put to elicit those strategies which they thought they would generally adopt when abstracting, and others were put to elicit those strategies they thought they would adopt when abstracting particular texts (see next section).

As for the source product data (section 7.2.2 above), because of the sheer bulk of material contained in the source texts, the selection of informants to furnish opinions on their successfulness was practically limited to those people who had already had to read them as a part of the exercise; otherwise informants would have been involved in massive amounts of extra work. Therefore the teachers were considered the most appropriate candidates. In addition, the abstract writers could have been asked to judge the source material, but the fact that each student only read two of the eight source texts at very most would have made comparing between texts problematic. Consequently only the teachers were asked to rate the success of the originals.

As for the abstracts process data (section 7.2.3 above), insights were elicited from both teachers and producers. The tripartite strategy was once again adopted, asking producers how they had selected elements from the originals,

how they had ordered them, and how they had realised the wordings. This was repeated for the teachers who were asked how they made these decisions generally when writing abstracts.

As for the abstracts product data (section 7.2.4 above), evaluations could have been demanded of both teachers and students. The teachers had already done this informally in class evaluation sessions (Chapter 6), but a more rigorous evaluation was called for. In fact the teachers were deliberately *not* asked to fulfil this role, since it was felt that they could be made better use of in the other three areas. Judges had to be chosen to represent the best possible model of real consumers of such abstracts. Ideally such judges had to be making use of such material as an everyday part of their working lives. Eight members of staff in BT Laboratories' main library were selected to judge the Information and Library Science abstracts. They were considered the most appropriate consumers for two reasons: firstly, as library professionals they all make extensive use of abstracts; secondly, they could be expected to be familiar with the subject matter discussed in the abstracts. Members of the Human Factors Division were selected to judge the General Knowledge abstracts. In addition to consumer judgements, the intuitions of the producers were also elicited, but in a slightly more focused way, asking them to identify the most/least successful sections of their abstracts. This was done to try to complement the gross, overall evaluations of the consumers, and also to try to pinpoint the specific difficulties which arose.

The next section concentrates on what was asked of the various informants.

7.3.2 Selecting the Most Appropriate Questions to Ask

In section 7.2 above, the discussion of the requirements was organised in terms of product versus process, and abstract versus original. In this section, on the other hand, the discussion will be organised in terms of the various questionnaires, each of which was administered to a different type of informant. In what follows, then, QJ1 was designed to elicit the opinions of model consumers of Information and Library Science abstracts; QJ2 those of the student abstract writers; QJ3 those of their teachers; and QJ4 those of model consumers of General Knowledge abstracts. The numerical ordering of these is deliberate, and reflects relative importance; hence QJ1 was designed to elicit the most important type of data, while QJ4 was designed to elicit the least

important type of data, the General Knowledge success judgments being collected for comparative purposes only.

This means that the different types of required data factored out in 7.2 above are combined in the questionnaires such that some informants are asked both process and product type questions about both source and abstract texts. Points of commonality and divergence between the various questionnaires are diagrammed below in 7.4, to provide a clearer comparative picture.

The questionnaires will now be discussed in turn.

• QJ1

The first judgements questionnaire, QJ1, is represented in Appendix 7.1, was designed to elicit the opinions of model consumers of Information and Library Science abstracts, and collected both qualitative and quantitative information (see 7.2.4 above). The following diagram shows that QJ1 was meant to gather only abstracts product data:

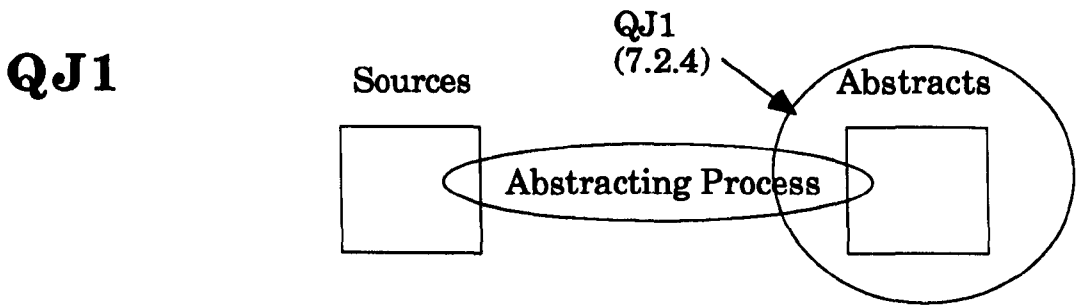


Figure 7.1: Schematic Diagram of QJ1

Each informant was given exactly the same task to perform: each was presented with four sets of abstracts, a 'set' being a collection of novices' attempts at the same original text, one set for each of the four different library/information science texts used. A QJ1 questionnaire was attached to each of the four sets.

The first section of QJ1 was labelled 'Instructions for this Questionnaire' and contained four paragraphs of rubric. Paragraphs 1, 2 and 4 are identical throughout all four versions¹ of the QJ1 questionnaires. For the sake of

¹There were four versions of QJ1, corresponding to the four different Information and Library Science source texts.

clarity, paragraph 3 explained what is specific to each version: it provided a full reference to the original text, gave an indication as to its type (in this case, whether journal article or book chapter), and stated how many abstracts the informant was expected to read and evaluate. This last number reappears in the second section as the number of slots provided in the box in which informants were asked to write their rankings. The fourth paragraph contained the instructions to rank the abstracts, and was as follows:

What I would like you to do is to RANK these abstracts according to how helpful you believe them to be. Please write the code letter which appears below each abstract next to its appropriate ranking. For example, you might consider C's attempt to be the most helpful, A's to be decidedly unhelpful, with B's somewhere in between. In this case you would write:

1.	C
2.	B
3.	A

This represents the quantitative part of the questionnaire. The first sheet of QJ1 ended with the instructions 'please use the space on the back of this sheet to indicate reasons for your ordering' and 'Please turn over' so as to direct informants' attention to the qualitative part of the questionnaire, which asked them to provide reasons for their rankings.

• QJ2

The second judgements questionnaire, QJ2, is represented in Appendices 7.2 and 7.3, was designed to elicit the opinions of the student abstract writers, and, although exclusively qualitative in nature, was altogether more complicated than QJ1, as the following diagram shows:

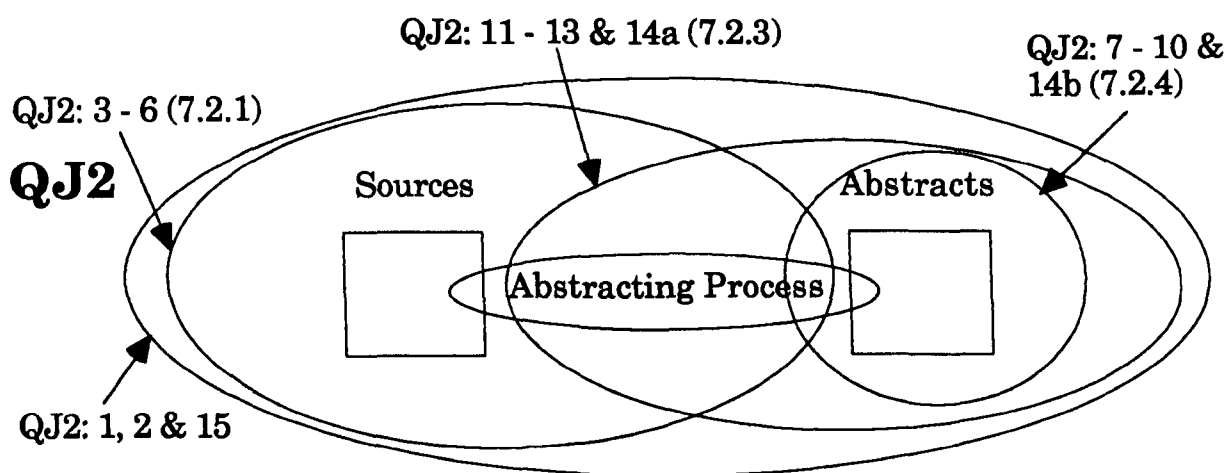


Figure 7.2: Schematic Diagram of QJ2

There were two versions of QJ2, but these are only minimally different. QJ2(1), represented in Appendix 7.2, was designed for those 16 students who wrote one abstract as part of the experiment, while QJ2(2), represented in Appendix 7.3, was given to the 13 students who wrote two. The questions in each are extremely similar and sometimes identical, the only difference being that in QJ2(2) the words 'abstract' and 'original' are made plural.

QJ2 contained fifteen questions, all of which were open and unstructured, as they were designed to elicit opinions rather than more directly comparable, quantifiable data. QJ2 starts with two general questions:

1. How difficult did you find writing the abstract?
2. How successful do you think you were?

These were included at the beginning of the questionnaire because it was thought that informants respond particularly well to questions which most centrally concern their own feelings or situation.

Questions 3 - 6 addressed the issue of the original text's 'summarisability', and so concentrated on data type 7.2.1, source material process:

3. Which bits of the original did you find hardest to summarise?
4. Can you give reasons why you think this was so?
5. Which bits of the original did you find easiest to summarise?
6. Can you give reasons why you think this was so?

Page 3 of QJ2, containing questions 7 - 10, mirrored QJ1 in that the way the questions were worded primarily concerned the producers' own abstracts viewed as product data (7.2.4), but, unlike QJ1, questions 7 - 10 of QJ2 asked informants to isolate those sections of their abstracts they were most/least happy with:

7. Which bits of your abstract do you think are most successful?
8. Can you give reasons why you think this is so?
9. Which bits of your abstract do you think are least successful?
10. Can you give reasons why you think this is so?

Questions 11 - 13 more generally concerned the abstract writing process, and only implicitly involve the original and abstract texts. The three questions were worded similarly, and sought to elicit information on the methods our informants (thought they) employed while writing their abstracts:

11. When writing your abstract, what strategies did you adopt when deciding what information to include?
12. When writing your abstract, what strategies did you adopt when deciding upon the order in which to present the information?
13. When writing your abstract, what strategies did you adopt when deciding how to express the information?

Such procedural knowledge was thought to be of enormous benefit, assuming it could be expressed at the appropriate level of specificity.

Question 14 was an attempt to offset the disadvantages of using non-expert (or, more accurately, non-practised) abstractors:

14. Suppose you had to do the exercise again. What changes do you think you would make a) to the strategies you adopted while writing your abstract? b) to the actual abstract itself?

Ideally, it would have been instructive to carry out a longitudinal study to trace the development of the students' abstracting competence; in this way it would have been possible to investigate the successful/unsuccessful continuum not just inter-producer (as has been done in fact) but also intra-producer over time. Unfortunately this was not a feasible option owing to the amount of time

allotted for the abstracting activity in the sample field (Chapter 6 above). Question 14 was thought to be a potentially useful compromise as informants were asked to hypothesise how they would perform facing the same task having had the experience of writing their practice abstracts (for the first time in virtually all of the cases), and having had these evaluated and discussed by their more experienced instructors.

Following the advice of much of the literature on questionnaire design, the final question, question 15, was the 'catch-all' question, which was an open invitation to the students to communicate to the researcher anything which they thought was worth recording:

15. Is there anything else you think I should know about your work, or your feelings towards it?

As with questions 1 and 2, this could have concerned any aspect of the exercise and its effect on the participants. This emphasis on the interpersonal nature of the task, rather than simply on its purely intellectual, problem-solving character, was considered particularly appropriate to realise the last question of QJ2.

- QJ3

The third means of collecting judgements, QJ3, is represented in Appendices 7.4 and 7.5, and takes the form of two very different questionnaires: a number of 'QJ3: Slips' and 'QJ3: General' (originally, each questionnaire had a single, small piece of paper attached on which informants were asked to record their judgments, hence they were called 'slips'. This particular lexicalisation has been retained even though current versions stretch to multiple numbers of A4 sheets).

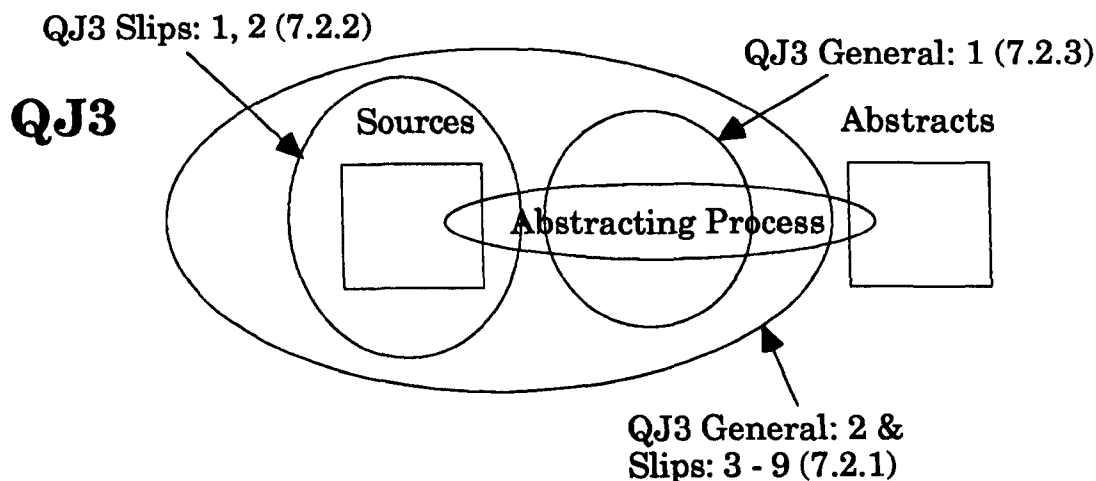


Figure 7.3: Schematic Diagram of QJ3

Both were administered to the two expert informants, the Information/Library Science teachers. QJ3 elicits both consumer- and producer-type judgements.

There are eight QJ3 Slips, one for each of the source texts used, together with one single QJ3: General, meaning that each expert was expected to complete a total of nine questionnaires. The slips will be discussed first.

- QJ3 Slips

There are eight slips, labelled 'QJ3: Slip for S1a' - 'QJ3: Slip for S4b'. The only variation between the eight is the wording for the source type in some of the questions, whether journal article, book chapter, or newspaper item. Tick boxes (questions 1 and 3) were preferred over the use of rankings because in this case the reasons underlying the opinions are sought for each and every text. Consequently the first question asked for an evaluation of the success of the original, while the second ensured that something was written concerning the reasons underlying the success or otherwise of every text:

1. How would you evaluate this < source text type, e.g. journal article > as an instance of its particular type?
(followed by five tick boxes, ranging from 'Very Successful' to 'Very Unsuccessful'; see Appendix)
2. Please give reasons for your judgement.

Notice that the rankings between the eight versions of QJ3: question 1 can be crudely compared to see which texts are preferred over others.

Questions 3 and 4 followed the format of 1 and 2, but asked about the 'summarisability' of the original:

3. How easy/difficult do you think it is to summarise this < source text type, e.g. journal article > ?
(followed by five tick boxes, ranging from 'Very Easy' to 'Very Difficult'; see Appendix)
4. Please give reasons for your judgement.

As with questions 3 - 10 of QJ2, an attempt was made to let informants focus the analysis through QJ3: questions 5 - 8. In fact, questions 5 - 8 are the direct equivalent of QJ2: questions 3 - 6, but are worth repeating to see if there are any interesting differences in the perceptions of informants who have had varying amounts of practice at abstracting:

5. Which bits of the < source text type, e.g. journal article > would you expect to be most difficult to summarise?
6. Please give reasons.
7. Which bits of the < source text type, e.g. journal article > would you expect to be easiest to summarise?
8. Please give reasons.

Question 9 follows the familiar 'catch-all' format, and invited informants to comment upon any aspect of the text, or its 'summarisability':

9. Do you have any further comments on the < source text type, e.g. journal article >, or its 'summarisability'?

- QJ3 General

The first question of QJ3 General was split into three, and exactly parallels the strategy questions of QJ2 (questions 11 - 13), pertaining to obligatoriness, order and realisation, respectively. Whereas in QJ2, students reported on their actual performance, the teachers were asked about the strategies they generally apply when writing abstracts:

- 1a. When writing abstracts yourself, what strategies would you adopt when deciding what information to include?
- 1b. When writing abstracts yourself, what strategies would you adopt when deciding upon the order in which to present the information?
- 1c. When writing abstracts yourself, what strategies would you adopt when deciding how to express the information?

Once again, these questions were focused through the inclusion of QJ3: Question 2, in which the teachers were invited to speculate how these general principles would relate to the eight source texts abstracted by their students:

- 2a. Do you perceive any differences between the eight source texts from the point of view of the strategies you would adopt for deciding what information to include?
- 2b. Do you perceive any differences between the eight source texts from the point of view of the strategies you would adopt for deciding upon the order in which to present the information?
- 2c. Do you perceive any differences between the eight source texts from the point of view of the strategies you would adopt for deciding how to express the information?

- QJ4

The fourth and last judgements questionnaire, QJ4, is represented in Appendix 7.6, and is identical to QJ1, saving some minor changes in rubric. Once again it involves abstracts viewed primarily as product data, and elicits judgements of success relating to the abstracts of the four General Knowledge source texts.

QJ4

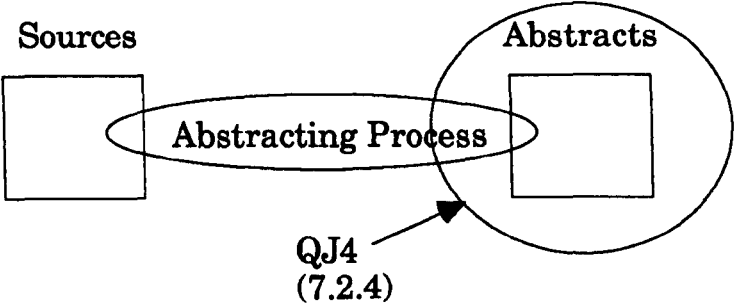


Figure 7.4: Schematic Diagram of QJ4

This was done for comparative purposes, since it was considered potentially advantageous to be able to investigate, for example, the degree to which the subject matter of the original, its specificity, the abstractor’s familiarity with the subject, or the match between the presuppositions of the abstractor and those of the reader, affect the writing of abstracts.

7.3.3 Selecting the Most Appropriate Layout and Means of Administration

Designing a questionnaire entails more than just phrasing questions carefully. A questionnaire that is confusing to look at can often distract informants such that they misinterpret what is asked of them, or, worse, fail to provide any answers at all. A well thought out graphical layout, on the other hand, can help encourage informants to respond in ways that will prove useful to the research.

Three basic goals were adopted in the graphic design of the various questionnaires. These were, in order of importance:

- 1) Questionnaires should be as clear as possible;
- 2) Questionnaires should be as short as possible;
- 3) Questionnaires should be as professional-looking as possible.

The first of these is generally recommended in the design literature. The second is desirable so as not to overly intimidate those respondents who are pressed for time, and it is thought that the third can help improve response rate (Youngman 1982: 20). Clearly there are trade-offs between these criteria -

occasionally the desire for brevity can run counter to the desire for clarity, for example - in those circumstances in which a compromise position could not be established, the order of the criteria was used to decide between competing design alternatives.

Youngman's recommendations for layout (1982: 20) were followed throughout. These include: the careful use of spacing; the inclusion of response boxes, which not only show respondents where to answer, but also gently hint at the length of answer expected; and, the clear differentiation between instructions and questions. Finally, a 'thank you' was placed at the bottom of each questionnaire, as well as in the covering letters (these being attached to those questionnaires which were not administered to informants personally).

Three points should be made concerning the presentation of the abstracts. Firstly, the students' handwritten versions were not used in their original form. Each script was manually typed into a computer, and misspellings regularised. Other orthographic features of the scripts, such as paragraph spacing, emphases, and so on, were maintained. Standard English spellings were used so as not to distract informants from the more interesting, and less well-understood, aspects of the investigation. Secondly, the typescripts were presented to the informants in a fixed order every time. This order was in fact that in which students handed in their work in the original experiment; so the abstract handed in first (though not necessarily the one that was finished first) became the one which was presented first (though not necessarily the one that was read first). The pros and cons of this ordering will be taken up in Chapter 8 below. Thirdly, each abstract was identified by letter only (A, B, C...). Authors' names were deliberately avoided, as other sociolinguistic studies have demonstrated that informants can be influenced by the sex or ethnic origin of speakers and writers. Informants were told little of the circumstances surrounding the preparation of the abstracts, and while a full reference was given for each source text, nothing was said about them.

7.3.4 Pilot Survey Data Collection

In all cases questionnaires were subjected to rigorous pilot testing. Two different kinds of people acted as pilot informants; they can be thought of as being either model senders or model receivers. The senders had themselves experience both of designing and of despatching questionnaires, being

professional Human Factors and Ergonomics researchers, two of them specialists in vision research. Receivers were selected from the staff of BT Laboratories' library. None of these pilot informants were used in further stages of the study. On the basis of their comments successive versions of questionnaires were generated, and retested. The final versions of the six different questionnaires used in this part of the research - QJ1, QJ2(1), QJ2(2), QJ3: Slip, QJ3: General, and QJ4 - are presented in the Appendices (see Appendices 7.1 - 7.6).

7.4 Conclusions

Designing good questionnaires is a time-consuming and difficult business. Bearing this in mind, and also the fact that a number of different types of survey data had to be collected, it is not surprising that a large amount effort was expended planning and administering this part of the data collection.

However, it was time very well spent. Most importantly, the fact that the abstracts product data was successfully collected meant that abstracts judged successful could be compared with abstracts judged less successful.

The opinions and judgements expressed by the various informants will be reviewed in the next chapter, Chapter 8. To conclude this chapter, Figure 7.5 overleaf collects together the various diagrams shown above in an attempt to make clear the scope of each questionnaire. The diagram also provides a means of comparing between questionnaires, to see which questions were shared between questionnaires and therefore informants, and which were unique to that particular questionnaire.

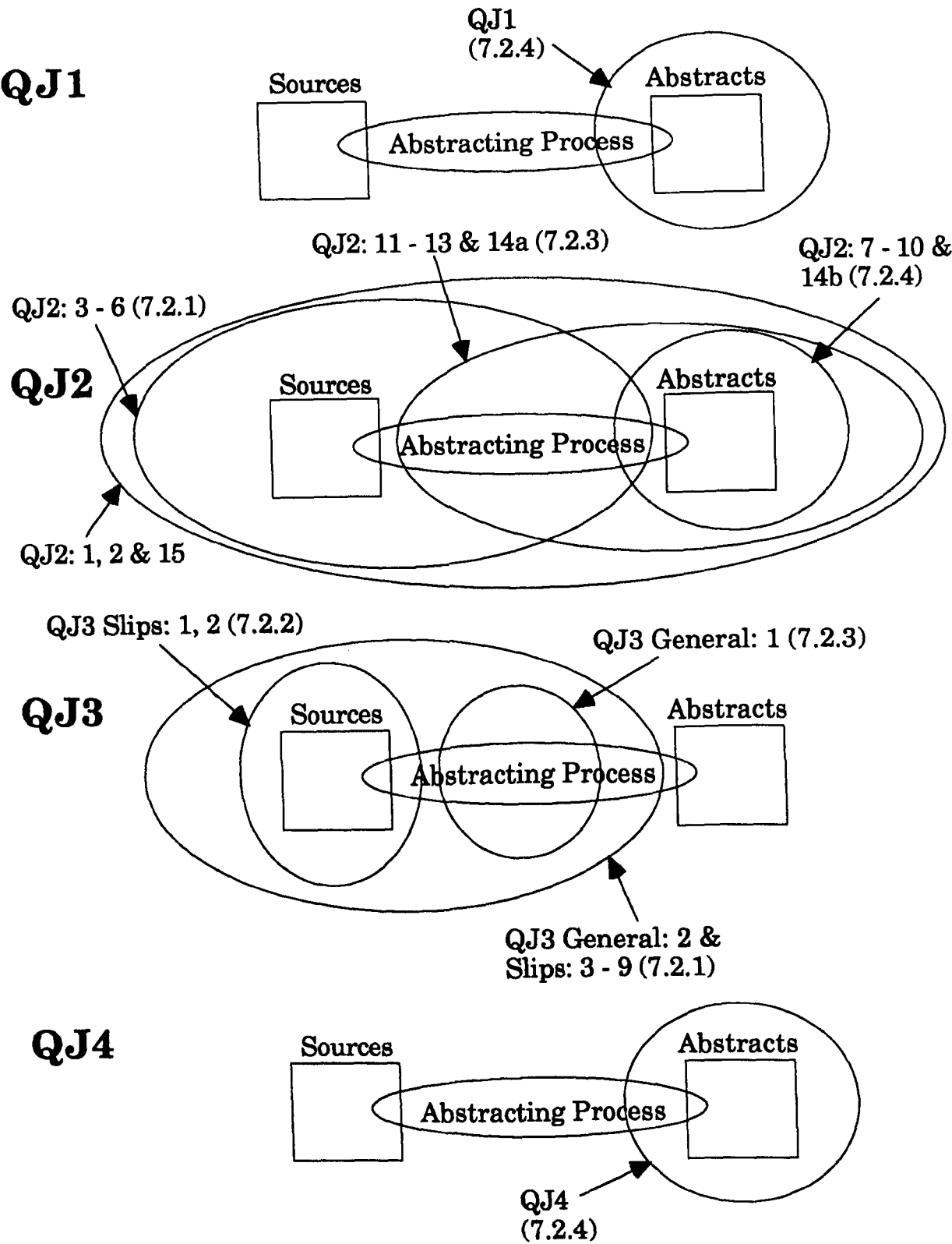


Figure 7.5: Schematic Diagram of the Four Questionnaires

8

Results of the Data Collection

8.1 Introduction

This chapter provides a catalogue of the data which was collected to supplement the naturally occurring abstracts discussed in Chapter 6 above. Most of this supplementary data was elicited by a number of questionnaires (Chapter 7), although reference is occasionally made to the lecturer - student interaction in the five lectures. This interaction can be thought of as being naturally occurring data (transcripts¹ of which appear in Appendix 8.5), though it must be acknowledged that the presence of the researcher will have had some effect on the proceedings.

There were four types of data collected by means of questionnaire: intuitions about source processes; judgements concerning the source material viewed as products; intuitions about the abstracting process; and judgements concerning the abstracts viewed as products. The results described below do not progress questionnaire by questionnaire; rather the discussion is organised to mirror the previous chapter, such that the requirement to capture the four different types of data (see sections 7.2.1 - 7.2.4 above) are treated in turn in sections 8.2 - 8.5 below. As will be seen, three of the four types of data were obtained from more than one source: source product judgements were collected from the lecturers only; both types of process data were collected from lecturers as well as students; and judgements concerning the abstracts were collected from the lecturers, their students, and from model consumers (8.5.1 - 8.5.3).

To return to a distinction first made in Chapter 1 above, the judgements made by consumers concerning the abstracts viewed as products can be thought of as being *external measures*. That is to say they constitute subjective

¹Conventional orthography is used to represent the transcriptions of the lecturer - student interaction in the Appendix; this is because it is the content of the discourse which is of interest, rather than its linguistic or phonological form.

assessments of the texts, and are driven by, but are logically separate from, certain linguistic and quasi-linguistic features of the texts. These features must be measured *internally*; that is to say, by means of linguistic apparatus capable of describing and articulating the particular forms and functions of the language which collectively determine the consumers' success judgements. This chapter can therefore be regarded as reporting on the collection of, among other things, external measures. Part Four, on the other hand, will investigate a number of different kinds of internal measure, and will report on how well each predicts the external measures already collected.

Each of the sections numbered 8.2 - 8.5 begins with a brief set of questions that phase of the data collection was designed to answer. The data collected is then discussed. The concluding part of each section then evaluates the extent to which the data collected successfully answers the original questions.

Finally section 8.6 provides a brief conclusion and paves the way for subsequent analysis.

8.2 Source Process Data

Source process data had to be collected as a first step towards answering the following types of question:

- What were the processes involved in the production of the original texts?;
- What were the processes involved in the interpretation of the original texts?;
- How does an original's 'abstractability' affect the process of writing an abstract?

The most suitable informants to answer the first of these questions were obviously the authors of the originals themselves, but it was not considered practical to collect data from such sources.

The most suitable informants to answer the second of these questions were firstly the students, and secondly, the lecturers. However, it was decided that matters concerning interpretation should be conflated with questions concerning the 'abstractability' of the original material. This was done for two reasons: firstly, to reduce the number of questions respondents were expected to answer; and secondly, because it was thought that probing for interpretative processes was altogether more problematic than probing for issues of abstractability, the latter being arguably less 'hidden' from introspection.

'Abstractability' perhaps deserves some brief discussion. Abstractability is assumed to be a property of source material, but it is a concept which lies at the boundary between, on the one hand, those processes involving the production and interpretation of the source, and those processes involving the production of the abstract, on the other. For present purposes, the abstractability of an original text is taken i) to be determined by certain of those processes involved in its production, and ii) to determine certain of those processes involved in the production of suitable abstracts.

For the purposes of the following discussion, this section will therefore consider the lecturer's response when asked to identify sections of the original material considered to be hypothetically more or less abstractable. Section 8.4 below will consider those processes involved during the actual writing of the abstracts.

What is particularly noticeable is that the lecturer has, in every case, pointed toward sections of the original, but in a way which does not permit their exact identification: 'a) Information transfer b) Information science' (Vickery and Vickery, Chapter 1, question 5); 'Trait studies' (Distinct Personality Type, question 5). The reasons underlying the judgements are perhaps more interesting: the reasons concerning those sections judged hardest to summarise principally involve matters of information content, and the decision whether to include or exclude certain facts in the abstract ('In the record of flooding a number of instances are documented and it is necessary to decide to what extent, if at all, they should be enumerated' (Erosion), 'The number of figures cited is quite large' (Cocaine)); whereas the reasons concerning those sections judged easiest to summarise principally involve matters of style, structure, or the development of the argument ('The largely negative stereotypes are ably reviewed ... Conclusion and recommendation are clearly established' (Distinct Personality Type, question 8), 'Sections are fairly short and the argument is easy to follow' (Vickery and Vickery, Chapter 2, question 8)).

Given the aims and objectives of this part of the data collection, to what extent did the actual data collected successfully answer the questions stated at the beginning of this section? This phase of the collection can only be considered to have been partially successful. There were two problems which hampered understanding: firstly, the inaccessibility of the authors of the source material; and secondly, the fact that such processes are notoriously difficult to collect from even the most willing of informants. Fortunately, the remaining three types of data were collected altogether more successfully.

8.3 Source Product Data

Source product data had to be collected as a first step towards answering the following types of question:

- Given that the success of an original text will affect the ease with which it may be abstracted, how successful was each original perceived to be?;
- Were all the original texts perceived to be equally successful, for example?;
- If there were different levels of perceived success, what reasons did judges provide for preferring some over others?.

The most suitable informants to answer questions such as these were the lecturers themselves; they were the only people who could be expected to have read all the source texts without being specifically asked to do so by the researcher. Data concerning the success of the source material were elicited using questions 1 and 2 of the third judgements questionnaire, QJ3 Slips (see Figure 7.5 above). These questions were as follows:

1. How would you evaluate this < source text type, e.g. journal article > as an instance of its particular type?
(followed by five tick boxes, ranging from 'Very Successful' to 'Very Unsuccessful'; see the Appendix to Chapter 7)
2. Please give reasons for your judgement.

The slips for the third questionnaire were administered to the two lecturers, but *only one was returned*. For the General Knowledge texts, the answers to question 1 were as follows:

Very Successful:	Erosion; Nature/Nurture
Successful:	Brain Drain
Moderately Successful:	Cocaine
Not Very Successful:	nil
Very Unsuccessful:	nil

For the Information Science texts, the answers to question 1 were as follows:

Very Successful:	nil
Successful:	Distinct Personality Type; Tanzania; Vickery and Vickery, Chapter 2
Moderately Successful:	Vickery and Vickery, Chapter 1
Not Very Successful:	nil
Very Unsuccessful:	nil

The reasons given for these judgements were elicited using question 2 ('Please give reasons for your judgement'); the answers provided are varied, but remarks involving *structure* frequently recur. For example, the Erosion article is applauded for its clarity, 'The narrative falls into four distinct sections, plus a conclusion', and the fact that the informant adds 'The inclusion of headings to identify the sections would enhance presentation' presumably means that the structure is apparent in the writing itself, rather than by virtue of its layout. The Cocaine article, on the other hand, one of the two originals considered least successful, although having a recognisable structure, was criticised as being 'rather discursive and information on a topic is sometimes scattered ... The headings used to label sections tend to be "catchy" rather than informative ... and each section covers a miscellaneous assortment of facts in many cases which are only tenuously connected'.

Turning now to the Information Science source texts, the remarks above are echoed by those pertaining to Vickery and Vickery, Chapter 1, whose 'discursive' nature is also noted, and explained by its being an introductory chapter; the same lecturer says to the students 'Chapter One in some ways was more difficult because it is just simply really an introduction to the book, and his aim in writing the chapter is, as I say, is, first of all, introductory, and secondly, a brief historical look at the way in which information science has developed' (Appendix 8.5, lines 1401 - 1405). Returning to the questionnaires, the Tanzania article is thought to be 'not difficult to understand and the argument is developed systematically', while the informant says of the Distinct Personality Type article 'The argument in this article is ably developed and organised under headings which indicate adequately the content of the various sections'.

Given the aims and objectives of this part of the data collection, to what extent did the actual data collected successfully answer the questions stated at the beginning of this section? This phase of the collection can be considered to have been reasonably successful. The General Knowledge source material was considered to be more successful overall than the Information Science source

material, but it is interesting to note a greater perceived variation in the General Knowledge originals. Of the four Information Science originals, three were thought to be similarly successful, with Vickery and Vickery, Chapter 1 'moderately successful'. In terms of the reasons given, the qualitative data collected was especially interesting; this informant was particularly sensitive to structure and to typology more generally (see also the next section below), and has provided information at a level of specificity far beyond that normally seen in the abstracting literature.

The next two sections describe the data collected which concerns not the source texts, but the abstracts: 8.4 discusses abstracts process data, and 8.5 below discusses abstracts product data.

8.4 Abstracts Process Data

Abstracts process data had to be collected as a first step towards answering the following types of question:

- How do writers select information from source texts for inclusion in their abstracts?;
- How do writers order this information in their abstracts?;
- How do writers put this information into words?.

Since abstracts process data is mainly concerned with the strategies that abstractors adopt while writing their abstracts, this type of data was collected by questionnaires administered both to the two lecturers, who were experienced abstractors themselves, and to all the students. In addition, the written guidelines that the students were given immediately prior to their practical exercise are relevant and are reproduced as Appendix 8.1: these were adapted by AS from recommended procedures published internally by the Department of the Environment. The guidelines are similar to those discussed in Chapter 5 above, and are evaluated by AS himself in the first session (see Appendix 8.5: lines 25 - 113).

The students' responses will now be considered, followed by an assessment of the lecturer's answers.

8.4.1 Students' Intuitions

QJ2, questions 11 - 13 and 14a elicit information about the strategies that the students employed in the writing of their abstracts: 11 - 13 concern obligatoriness, ordering and realisation, respectively; 14a was designed to partly compensate for the fact that the subjects cannot be considered experienced abstractors, and asked them how they thought their strategies might change with more practice.

When asked how they decided what information to include in their abstracts, most of the fourteen respondents signalled that they read the original a number of times, took notes in some cases, and then chose the ideas which seemed most important to them. Typical replies were: 'Firstly I read the article through a couple of times to get a general idea of the subject matter. Then I began to write the abstract by noting down and summarising the salient points in each paragraph or section'; 'I firstly read through the original, marking main points and areas that I wanted to include in my abstract. If the article was quite lengthy or I hadn't understood it in places, I found that reading parts of it again gave me a clearer idea of what information I had to include in my abstract'. Because most of the responses to question 11 indicate that the major strategy adopted was to include that which the abstractors considered most important, with the benefit of hindsight it may have been useful to have also asked the question 'how did you know what was the most important material?'. It is interesting to note that only one student makes even the slightest reference to the perceived needs of the audience, 'I decided to include information on the trafficking of drugs, and read through the article looking for references to it. I left out statistics etc as I felt this was not needed in an abstract. Meant to be a general abstract', something which would have dismayed the lecturers who stressed the need to carefully consider the purpose and users of an abstract (Appendix 8.5: lines 28 - 39), and the possibility of 'audience slanting' (Appendix 8.5: lines 118 - 143). Interestingly, the most unusual intuition was 'I read through the articles several times, and then jotted down what I could from memory, on the (quite likely mistaken!) belief that these would be the most important points'; mistaken or not, this would delight cognitive psychologists who would take this remark as vindication of the type of experiments surveyed above, in which memory for text is often thought to be a means of deriving the 'gist' of the message (see Chapter 3).

When asked how they decided upon the order in which to present the

information in their abstracts, ten of the students said they followed that of the original. Most of the others seem to have used the original order as a template, but have included sections in their own abstracts which brought together material from different locations in the source material: 'Decided introduction should state what the article overall is about. Then should mention major points raised eg stereotypes, and then conclude with author's conclusion'; and 'In some ways followed the pattern of the original when abstracting the information. But really I wanted to sort of encapsulate all my information in the end in one or two paragraphs, which ran smoothly into one another, but which did not particularly follow the pattern of the article'. One student's reply, 'None. I merely translated it down through the hierarchy with which my brain disseminated it. Almost an unconscious kind of dissemination', neatly demonstrates how difficult it is trying to gain access to processes which are not available to intuition, although he does go on to say 'I notice, I mentioned student and higher education before employer. This may have been because the article was headed "Graduate Careers" and was from the "Education" section. Had it been headed "Business", I may have ordered it differently (who knows!)', again, one of the very few references made to the perceived requirements of the user.

When asked how they decided how to express the information in their abstracts, replies were divided: some intimate unconscious 'guesswork'; some the desire to produce a text that was easy to understand, as simple as possible, and as concise as possible; some used the language of the original ('Followed the language of the original. Linked the major ideas, themes, terms with the least and simplest language I could'; and 'My abstract is composed of sentences/words taken directly from article and put together to form comprehensible comments. None of it includes my own words, except to tie up sentences etc!'), while others suggest more extensive editing and invention ('Above all I tried to totally avoid pronouns and adjectives and just to list the ideas contained in a text in almost a list, without conjunctions or anything to try and improve the style of prose. Basic'; and 'I wanted the first sentence to be eye-catching, like the first few sentences of a novel, so that the reader would want to find out more ... I also wanted to target the article to the sort of people I thought would want to read it, i.e. research students').

When asked how they thought their strategies might change with more practice (question 14a), most said that they would employ essentially the same strategies. One informant's reply captures the general mood well: 'Difficult without subsequent training/teaching. I approached it instinctively and think

I would do so again'. This last comment nicely highlights one of the primary motivations of this work: to make more explicit those strategies which can aid the production of successful abstracts.

8.4.2 The Lecturer's Intuitions

The data elicited from the lecturers by means of QJ3 General is on the whole more specific and positive than that gleaned from the student producers. The lecturer's answers to question 1 are comprehensive and so are reproduced here verbatim; this is his answer to question 1a:

(A) Articles.

(a) Scan the text, paying particular attention to the introduction and conclusion. Note any section likely to pose problems in summarisation.

(b) Selection of information for inclusion varies with the nature of the original: for example:

1) Result of empirical research. Convey the hypothesis, the conclusion and provide key supporting argument and measurements.

2) Descriptive article. Identify the parts of the narrative and decide whether any part is sufficiently important to warrant extra attention or if a balanced treatment of content will suffice.

3) Theoretical article. Identify the hypothesis put forward and indicate the nature of the arguments for and against it. In social sciences it is often necessary to state the author's school of thought and the methodology employed. In non-sciences the quality of the argument is often a major factor and this raises the question of evaluation in abstracts, not, of course, normally found.

4) Review article. One of the more difficult types to deal with as it entails summarisation of material which has already been summarised and the argument of a source may not have been dealt with adequately. Best handled perhaps by providing an indicative abstract; if not, then it is necessary to convey the main findings of the original sources, with any data which are essential to the argument.

(B) Books.

Read the introduction and conclusions carefully; examine the contents list and scan any chapter(s) which summarise the findings of any part(s) of the work.

It is interesting to note that the lecturer considers there to be four major genres of article, and that each warrants a slightly different strategy for deciding what information to include, and, as can be seen in the answer to 1b,

for deciding on the order in which to present the information:

1. Where the writer is putting forward an hypothesis, developing a theory, etc. this is first stated and followed by conclusions, if any. This ensures that the reader is provided with essential information immediately - especially important where the abstract is lengthy. Supporting argument, etc. will follow, usually in the order of the original text, but bringing together any related ideas scattered in the narrative.
2. For descriptive works, the subject is first defined - unless this is fully conveyed by the title - thereafter the order of presentation in the original is satisfactory. Again, however, if related information is dispersed it will be necessary to restructure the narrative.
3. Review articles are often a problem. The article is often arranged by texts surveyed by the author and these sources are usually putting forward arguments on the same issues, so it is desirable to structure the summary by topic, point of view, etc. rather than adopt the original pattern.
4. For order of elements within an abstract I prefer: Citation - abstract - statement of illustrations, bibliography, etc. in the original.

The somewhat shorter answer to 1c, however, would seem to suggest that the strategies for deciding how to express the information are a good deal less dependent upon source type:

1. Prepare a draft for the abstract and eliminate any unnecessary terms. Ensure that the text follows rules laid down for use of tense, mood, etc.
2. Consistency in usage is paramount; I follow the rules most frequently suggested. Use direct statement rather than indirect; present tense for conclusions, generalizations and content, with past tense for work done and observations made before writing the article being abstracted.

Turning now to question 2, which focuses on how the particular items of source material might affect the three kinds of strategy adopted, 2a is especially useful in that it conveys very good clues as to which genre each source belongs:

1. Items 1a and 1b illustrate the points made under "descriptive article" in para 1a above.
2. Items 2a and 2b both cover the issues raised under "Review article" (para (b)4 in 1a above).
3. Item 3a is a research paper with results based on examination of case studies rather than measurement of variables; 3b is more straightforward as it consists largely of hypothesis and recommendations for an alternative policy.

4. Item 4a includes both a "theoretical" component (ie analysis of the concept information science) and descriptive features; item 4b is "descriptive" in content.

The lecturer views, therefore, the Erosion and Cocaine articles as primarily descriptive; the Brain Drain and Nature/Nurture articles as reviews; Distinct Personality Type presumably as an article embodying the 'result of empirical research'; Tanzania presumably as a 'theoretical article' (the difference between these last two types is sometimes a little difficult to see; notice that the four types given in 1a become three in 1b, with 'empirical' and 'theoretical' research perhaps being conflated); Vickery and Vickery, Chapter 1 as a hybrid of 'theory' and 'description'; and Vickery and Vickery, Chapter 2 as primarily descriptive. From these categorisations, and from the lecturers' oral assessments (for example, Distinct Personality Type is discussed in lines 1523 - 1557 of the transcript, Tanzania in lines 1497 - 1503, Vickery and Vickery, Chapter 1 in lines 1406 - 1440, and Vickery and Vickery, Chapter 2 in lines 1441 - 1460), it is possible to derive the information the lecturers believe should have been included for maximum effect. Similarly, question 2b provides pertinent advice on its ordering:

1. Item 3a [Distinct Personality Type] deal first with hypothesis / results / recommendations, followed by case studies.
2. In items 1b [Cocaine], 2a, 2b, [Brain Drain & Nature/Nurture] information pertaining to a specific topic(s) is scattered in the text and should be integrated.
3. In other cases [Erosion, Tanzania, Vickery and Vickery, Chapter 1 and Vickery and Vickery, Chapter 2] the order of the original may be followed.

The advice and comments so far provided are sufficiently specific to be of considerable use. The odd man out, as it were, is realisation. The lecturer's comments given in answer to 2c mention only tense:

The one exception is item 3a [Distinct Personality Type] in which the summary of the case studies will be in the past tense, the rest of the abstract in the present tense. In all other items the present tense is appropriate.

A consideration of the lecturers' oral assessments, the students' own remarks in QJ2, and the lecturers' opinions given throughout QJ3 General and in the Slips, reveals very little discussion of the actual language of the abstracts. The precise mechanisms underlying the way material has been (in the case of the

students), and should be (in the case of the lecturers), realised linguistically remains, then, largely hidden from view; this is in stark contrast to those opinions elicited from the consumers (8.5.3 below), whose remarks concerning such realisations are very much to the fore.

In spite of their not having written abstracts for the source material themselves, the lecturers were also asked to hypothesise 'How easy/difficult do you think it is to summarise this [newspaper item/journal article/book chapter] ?' (QJ3 Slips, question 3). The lecturer's response for the General Knowledge texts was as follows:

Very Easy:	Erosion
Easy:	Cocaine; Brain Drain; Nature/Nurture;
Moderately Easy:	nil
Difficult:	nil
Very Difficult:	nil

The lecturer's response for the Information Science texts was as follows:

Very Easy:	nil
Easy:	Distinct Personality Type; Tanzania; Vickery and Vickery, Chapter 2
Moderately Easy:	nil
Difficult:	Vickery and Vickery, Chapter 1
Very Difficult:	nil

On the whole, then, the Information Science texts are thought to be harder to summarise. The ordering of the two book chapters (Vickery and Vickery, Chapter 1 and Vickery and Vickery, Chapter 2) is corroborated by the same lecturer's comments in one of the oral evaluation sessions (Appendix 8.5: lines 1401 - 1405); the lecturer who was responsible for evaluating the student's abstracts of the Distinct Personality Type and Tanzania originals did not return the third judgements questionnaire, QJ3, but it is interesting to note that she differs slightly from the ranks given above, because she considers Distinct Personality Type to be 'much more difficult' to abstract than Tanzania (Appendix 8.5: line 1519).

There is a certain degree of similarity between the originals ranked according to abstractability and those according to success (see 8.3 below): Erosion, one of the two most successful originals, is also considered to be the

most easily abstractable; Nature/Nurture, the other item designated 'Very Successful' is not considered 'Very Easy' to summarise, but falls one place to 'Easy' to summarise; the four originals considered 'Successful' are among the six considered 'Easy' to summarise; and Vickery and Vickery, Chapter 1, one of the two least successful originals, is also thought to be 'Difficult' to summarise, the only one of the eight in fact to receive anything less than a 'Moderately Easy' to summarise marking. The reasons given for the ranks (QJ3 Slips, question 4) include: the quality of the writing, and the fact that 'the text is not cluttered with excessive data' (Erosion); 'Most of the problems stem from the structuring of the text' (Cocaine); 'The narrative is straightforward and the argument readily understood' (Brain Drain); 'The narrative develops logically and the parts of the argument are clearly distinguished' (Nature/Nurture); 'The argument is systematically developed and readily interpreted in most of the sections. Points may easily be summarised in these areas' (Distinct Personality Type); 'Section headings are few and inadequate as an aid to the identification of the article's main themes ... The combination of distinct themes in a single section (eg planned development and generation of funds) also hinders extraction of information' (Tanzania); 'Perhaps the most difficult of the titles to abstract owing to the disparate nature of the contents. An indicative abstract would present few difficulties, but the informative type demands differing approaches to the three types of argument employed' (Vickery and Vickery, Chapter 1); and 'Provides balanced treatment of a number of elements within the scope of the subject ... Therefore the abstractor is able to give equal weighting to each section of the narrative ... Section headings are, on the whole, helpful in the identification of topics' (Vickery and Vickery, Chapter 2). Clearly to this informant then, the structure of the original is an important determinant of a text's abstractability.

Given the aims and objectives of this part of the data collection, to what extent did the actual data collected successfully answer the questions stated at the beginning of section 8.4? Once again, this phase of the collection can be considered to have been reasonably successful, especially since it must be borne in mind that the processes to be investigated are not directly available to intuition; in effect, respondents are being asked to verbalise thought processes which could be argued to be at a semi-conscious level. Given this proviso, it is perhaps not surprising that some of the students' responses are less than fully precise. The lecturer, on the other hand, has provided guidelines, which, although still not always directly operationalisable, nevertheless are more useful than much of what is published in the normative literature (see

Chapter 5 above).

8.5 Abstracts Product Data

Abstracts product data had to be collected as a first step towards answering the following types of question:

- What reasons do readers give for preferring one abstract over another?;
- Do readers think alike? More specifically, to what extent do they agree with each other in their various preferences?;
- Is 'success' better explained by correlation with one, or with many, linguistic variables?.

Abstracts product data is the most important data type to be collected, since success judgements concerning the abstracts viewed as self-supporting texts in their own right are a necessary precondition for answering the basic research questions addressed in this thesis (the three questions above are a selection of those questions stated in Chapter 1 above).

Data concerning the success or otherwise of the various abstracts were collected from a number of different sources in a number of different ways: judgements questionnaires QJ1 and QJ4 (Chapter 7) elicited both qualitative and quantitative judgements from informants who represented genuine consumers of abstracts from the field of Information Science (QJ1), and from a General Knowledge perspective (QJ4); questions 7 - 10 and 14b of QJ2 (see Figure 7.5 above) were directed towards the abstract writers in an attempt to tap the intuitions of the producers, how successful they considered their own abstracts to be; in addition, audio transcripts were made of the lectures, two of which contain evaluative material (sessions 3a, 3b and 5 in Appendix 8.5). The three different sources of data will now be discussed.

8.5.1 Instructors' Opinions

Two of the five sessions that were tape-recorded (transcripts of which appear in Appendix 8.5) consisted of oral evaluations by the lecturers of the students' abstracts. The first of these, in which the class was split into two according to whether they had written an abstract from S1 or from S2, pertains

to the General Knowledge source material, the second to the Information Science.

The first evaluative session was led by lecturer SH with those students who had written General Knowledge abstracts for either the Erosion article from the 'Geographical Magazine', or for the Cocaine article from the 'Economist', or both. As mentioned above, the tape-recorded sessions were not controlled in any way by the researcher, and, as might be expected from the lecturer's initiating comments '... What I'd like to do now is to get some general discussion going on how you found writing the abstract, what your methodologies were, how you set about it' (Appendix 8.5: lines 670 - 672), the discussion is participative and mainly process-based. Individuals' abstracts are not discussed, the lecturer preferring (perhaps in deference to the tenor of the open class situation) to speak in general terms about strategies employed. Where product-type judgements do occur, these once again involve levels of detail (the lexicalisation 'bogged down in detail' appears again, Appendix 8.5, lines 713 - 714), specialist knowledge, and questions of audience (Appendix 8.5: lines 732 - 756).

At one point the lecturer asks 'What about the form of the abstract? A lot of you didn't look at Alex's notes in great detail. But did you generally find that the structure was coming out in a particular kind of way? In terms of introduction, and conclusion, and the body of the text. How did you set about putting a structure on it? Anybody like to say anything about that?' (Appendix 8.5: lines 794 - 798), which is an extremely promising question from the point of view of both process and product data. Unfortunately, in the immediately following response, the student has interpreted the 'text' as being the *original*, and the topic shifts away from discussion of the form of the abstracts proper.

The second evaluative session was led by lecturer AS with those students who had written General Knowledge abstracts for either the Brain Drain article from the 'Independent', or for the Cocaine article from the 'Independent', or both. Most of AS' comments elucidate the strategies he himself would have adopted to abstract the newspaper articles, the naturally occurring analogue of those data elicited by questionnaire discussed in section 8.3 above, and in 8.5 below.

Specific abstracts are mentioned only twice: a misunderstanding made by one person, whom the lecturer does not name, is used to illustrate the point that unclear sections of the original can lead to mistakes in the abstracts; and one person, who is identified this time, is applauded for a particularly perspicuous and economical rendering of a long section of the original

(Appendix 8.5: lines 557 - 568). In addition, AS also warns of the dangers of getting too 'bogged down in detail' (Appendix 8.5: lines 575 - 576).

The third evaluation session consisted of both lecturers discussing the Information Science abstracting task, SH discussing the two articles, Distinct Personality Type and Tanzania, and AS the two book chapters, Vickery and Vickery, Chapters 1 and 2. Once again, the discussion here will be limited to opinions of the abstracts viewed as products, of which there are relatively few. AS noted a tendency that some of the students had included too much information from the beginning of the source material, with the result that the abstracts were not well balanced informationally (Appendix 8.5: lines 1374 - 1381); and, although he is careful not to mention the name of the author, he warns against the use of too much metadiscourse, 'The one criticism I have here: in one case the verbage could have been cut down quite considerably. For example, periodically the abstractor says "in this part of the chapter", or "within the chapter", or "the chapter thus sets the tone", and so on. There's no need to keep on repeating this business about the chapter. We're well aware that the abstract is restricted to a particular chapter, and all that can be cut out' (Appendix 8.5: lines 1444 - 1449).

SH reinforces AS' points concerning weighting (Appendix 8.5: lines 1480 - 1482), concluding that some of her students had fallen into a similar trap. She notes that both the 'structure of the abstracts' and their 'subject coverage' were generally good (Appendix 8.5: lines 1486 - 1496), and that at least some of this may have been due to what she perceives to be as clear writing in the originals. She also makes the interesting observation that, in her opinion, the 'Tanzania abstracts were generally better than the Distinct Personality Type abstracts (Appendix 8.5: lines 1497 and 1519). With regard to importing material wholesale from the originals, she says 'Some people did actually just sort of lift particular sections, changing the odd word here or there, which isn't a copyright problem. But it's tempting to do that when you see a particular sentence that just fits what you want to say. A little bit of the verbiage as well; which is a problem, you need to cut down on the, toppings and tailings really, and go straight for the nitty-gritty' (Appendix 8.5: lines 1509 - 1514); this would seem to suggest that such 'lifting' is permissible, providing alterations are made to ensure that such material is economically expressed and fits in satisfactorily with the co-text.

8.5.2 Producers' Opinions

Questions 7 - 10 of the second judgements questionnaire, QJ2, were designed to supplement the data discussed in the previous section by asking the writers of the abstracts to single out those sections of their work they themselves considered most/least successful, and to provide reasons for their various selections. This was done primarily to focus the analyst's attention in a grounded way, rather than appealing to his own intuition. For reasons discussed in the last chapter, question 14b was included to elicit the producers' opinions of how they would change their abstracts with the benefit of hindsight, as it were.

Fourteen questionnaires were received from the students, as already mentioned; these comments pertain to both the Information Science and to the General Knowledge abstracts. Of these fourteen, approximately half provide clues as to the location of the most successful sections of their abstracts. Five students prefer their beginning sentences: 'the initial couple of sentences seem fairly sound'; 'I think the first two sentences of my abstract are the most successful'; 'the first part'; 'the first paragraph'; and 'I think my introduction is quite interest catching'. The reasons given include levels of detail once again, conciseness, and the degree to which the abstractors understood that which they were trying to summarise. It appears that most of the students appreciate the importance of the first sentence of the abstract, the so-called 'topic sentence' in the abstracting guidelines, which might be thought of as being the meta-abstract, and that most felt that they had made a reasonable attempt at its construction. Many of the students were aware of having become 'bogged down in detail' shortly thereafter.

Although other respondents did hint at other sections, these were not always clearly particularised: one student wrote that the bits she thought were most successful were those which were not 'copied straight from the original', arguing 'I have understood these parts in the original the most, so my understanding is reflected in the abstract'; the least successful sections, according to another abstractor, included 'the handling of the more theoretical parts and those areas that were highly specialist'. The reasons given, probed by questions 8 and 10, invariably concern detail, length, subject familiarity, or the decision whether or not to include certain information from the original; comments concerning the quality of the students' own writing do not appear. Most of the answers to question 14b concern length, and the inclusion or deletion of particular points of information.

Many of these remarks are interesting, even though few of the students have responded as anticipated. Perhaps a better way to have identified particular sections of the abstracts would have been to have asked the producers to underline those sentences they considered most/least successful. This more structured approach would at least have ensured that more data were particularised, although it must be acknowledged that even this strategy has its disadvantages. The student's intuition that the best sections of her abstracts were those which were not imported wholesale from the originals is an interesting one, and one which has far-reaching implications for the design of automatic text summarisation systems (Chapter 4); this would not necessarily have emerged, however, had the underlining protocol been employed, it being an altogether less 'open' strategy. This demonstrates that questionnaire design is very much a 'swings and roundabouts' affair, a situation which vindicates the multimethod idea of not putting all one's eggs in one basket.

8.5.3 Consumers' Opinions

The last set of abstracts product data is the richest of the three; it is the only one in which each and every abstract receives both a rank (8.5.3.1), which enables comparisons to be made between different versions of abstract written for the same original material, and supporting comments (8.5.3.2), which were elicited to help the analyst home in on the linguistic features in the abstracts which might best explain the informants' orderings.

8.5.3.1 Consumers' Ranks

The following discussion will briefly consider the ranks, this data being made the subject of a small amount of statistical analysis. The consumers' ranks for the different abstracts of the eight source texts are shown in Appendix 8.4. In Set 1, for example, there were five abstracts, labelled A - E, written for the original article 'Is the librarian a distinct personality type?'. As can be seen, the first of the eight judges considered D to be the most helpful (denoted by the figure 1), and A the least (denoted by the figure 5). In fact, all eight judges consider A the least helpful; there is, however, considerably less agreement in other sets. A statistical measure designed to indicate to what

extent several judges agree or disagree is the Kendall coefficient of concordance, w (Siegel 1956: 229 - 238), and has been calculated for each of the eight texts. The results for the General Knowledge texts with values of w given to three decimal places are as follows:

	S	W
Erosion	86	0.137
Cocaine	408	0.405
Brain Drain	100	0.159
Nature/Nurture	288	0.457

The results for the Information Science texts are as follows:

	S	W
Distinct Personality Type	382	0.597
Tanzania	154	0.241
Vickery and Vickery, Chapter 1	170	0.531
Vickery and Vickery, Chapter 2	14	0.109

The calculation has been derived such that a w value of 0 means that judges are being completely random in their allocation of rank scores, whereas a value of 1 signifies complete agreement. Because the number of cases is relatively small, care must be taken not to over-interpret these figures. What is clear, however, is that there is a wide range of values, from a comparatively large amount of agreement (Distinct Personality Type), to almost complete disagreement (Vickery and Vickery, Chapter 2). The fact that there is such a wide range suggests two things: firstly, 'helpfulness' as quantified by these questionnaires appears to be a multi-faceted notion; and, secondly, if there are indeed components of helpfulness, judges may well differ over the relative importance they attach to them.

If valid, these last two points place a serious limit on the usefulness of collecting solely quantitative data. Informants' qualitative opinions will now be explored, therefore, to discover, firstly, what reasons informants give for preferring some abstracts over others, and, secondly, to what extent these reasons are shared between informants.

8.5.3.2 Consumers' Qualitative Remarks

The informants' opinions are presented verbatim in Appendix 8.4. These opinions are also categorised, so that similar opinions can be clustered, clear differentiation between different types of opinion can be made, and an overall picture of the classification can be given. This categorisation has been carried out by means of certain coding protocols, which are used to provide a bird's eye view of the mass of qualitative opinion; unlike much of the coding practised on qualitative data analysis (Miles and Huberman provide a useful appraisal of the strengths and weaknesses inherent in codes and coding, 1984: 54 - 64) which seeks to reduce large quantities of data in the form of words to something more amenable to numeric interpretation, the distribution of the codes is used primarily as a presentational device - to give an overall indication of the nature and scope of informants' verbal responses. Such qualitative data will be used to direct the pilot analysis discussed in subsequent chapters. It is important to further stress that the codes are merely a means of appreciating the entire view, before focusing upon particular details. As such, codes reduce the risk of being overwhelmed by the sheer bulk of the qualitative data, or of 'not being able to see the wood for the trees'. Following Miles and Huberman's advice (1984: 54 - 64), a list of codes was drawn up, and a sheet of 'operational definitions' produced. Codes were generated once the entire corpus of judges' remarks had been read many times, and were developed inductively (Miles and Huberman 1984: 57), preferring 'in vivo codes' over theoretical constructs (Strauss 1987: 33 - 34); in other words, only terms that would be meaningful to the judges were abstracted from their comments to become codes, concepts familiar only to linguists (such as thematic progression, for example) do not therefore appear. Altogether thirteen codes were used (see Appendix 8.2). Such a taxonomy is, of course, highly subjective, so the researcher's own codes were compared and contrasted with those of two other researchers² enlisted to help. The multi-coding process was conducted relatively informally (the instructions given to the other coders appear as Appendix 8.3), and, on the basis of the three sets of codes and discussion thereof, a set of codes was derived which best represents the collective view. Appendix 8.4 shows the qualitative opinions annotated with codes together with the quantitative rankings.

The chief benefit of having data coded in such a way (especially when stored

²Sandra Williams and Dawn Wright, the two other members of BT's automatic text summarisation project team.

as a database on a computer) is that it becomes far easier to manipulate. Correspondences are more easily charted, agreements and disagreements can be flagged, and even quite rough and ready counting operations can reveal interesting patterns in the data. For example, the distribution of codes in the Information Science data is as follows:

CONT	=	64
LAYO	=	9
LENG	=	44
MISC	=	5
PARA	=	9
PUNC	=	3
READ	=	66
REFS	=	4
STRUC	=	35
STYL	=	18
SYNT	=	10
TYPO	=	7

It is worth reinforcing the point that judges have been given a relatively free hand, as it were; the probe used to elicit these opinions was simply that of 'helpfulness'. These figures represent, then, an indication of the relative importance of the different features in the data the informants are either applauding or condemning.

It is perhaps not surprising that CONT, signifying information content, figures prominently. Because the judges are supposed to model real-world users of abstracts, they were given the abstracts without sight of the originals. Naturally a good deal of their comments therefore concern the relative amounts of detail and information furnished by the abstracts. It is clear from the qualitative data that it is important to strike a careful balance between including too little information and too much. For example, Judge 8 thought abstract version A written for the Distinct Personality Type original,

'too long, with too much individual detail'.

Similarly, Judge 7 considered abstract version C written for the Tanzania original, to contain

'too many individual details'.

Judge 4, on the other hand, wrote the following in response to abstract version C written for the second chapter of the Vickery and Vickery book:

'gives the impression of leaving out information that might be useful, although I like the style'.

Unfortunately it is less clear how these observations could be accounted for linguistically, especially since the decision was taken at an early stage in the research not to consider the originals in any great detail (Chapter 1).

LENG, scoring 44, seems to be another good predictor of informant satisfaction. Some abstracts are considered too short:

'is brief but at the expense of telling you about the paper itself'
(Distinct Personality Type, version E, judge 1);

'as for D but not quite as good - possibly just a bit too short'
(Distinct Personality Type, version E, judge 2);

'too short...'
(Distinct Personality Type, version E, judge 6).

In contrast, all the judges thought that Distinct Personality Type abstract version A was too long. Length is something that is debated thoroughly in the abstracting literature; the same is true for many of the comments coded as being LAYO, PARA, PUNC, REFS and TYPO. The following comments give a flavour of the types of criticism levelled:

'uppercase irritating'
(Distinct Personality Type, version A, judge 4);

'the citations tended to distract from the author's paper itself. No need for citations, I feel'
(Distinct Personality Type, version D, judge 6);

'"A" contains needless repetition that obscures the meaning'
(Tanzania, version A, judge 8).

These types of criticism should be relatively simple to correct. However, the remedy is much less obvious for those abstracts diagnosed as follows:

'very awkward style, clumsy ...'

(Distinct Personality Type, version C, judge 7);

'C and D are a bit waffley'

(Distinct Personality Type, versions C and D, judge 4);

'C and D failed to catch the attention and appeared bland and boring'

(Distinct Personality Type, versions C and D, judge 5);

'incoherent'

(Tanzania, version A, judge 4);

'very jerky portrayal of concepts. Didn't flow at all'

(Tanzania, version A, judge 6);

'I didn't like the style of D very much - too chatty possibly'

(Tanzania, version D, judge 3);

'fragmented and uncohesive'

(Tanzania, version D, judge 8);

'is totally hopeless - the layout is confusing and has no real structure'

(Vickery and Vickery, Chapter 1, version A, judge 1);

'unclear and difficult to follow what the writer is trying to say'

(Vickery and Vickery, Chapter 1, version C, judge 2);

'readability marred by rambling sentences'

(Vickery and Vickery, Chapter 1, version D, judge 8);

'disjointed'

(Vickery and Vickery, Chapter 2, version B, judges 3 and 4);

'a rambling, incoherent account, giving little useful information'
(Vickery and Vickery, Chapter 2, version B, judge 7).

It is precisely these types of comment, many of which are coded READ (signifying readability, also glossed as coherence in the code definitions), which are of primary interest, in that i) they feature prominently, and ii) they are not easily acted upon.

Given the aims and objectives of this part of the data collection, to what extent did the actual data collected successfully answer the questions stated at the beginning of section 8.5? The instructors' opinions (8.5.1) were understandably a little disappointing, since they were derived from discussion which was participative, process-based, and, because the evaluation was given in class and was not formally assessed in any way, programmatic. This meant that additional, elicited data had to be collected. The producers' opinions (8.5.2) were more instructive, although the focusing strategy, in which students were asked to point to particular sections of their work to help the analyst focus his attention, did not turn out as well as expected. Collection of the consumers' opinions (8.5.3), on the other hand, can be considered to have been extremely successful. This was particularly gratifying, since it was planned that this phase of the data collection would drive the application of the linguistics in the analysis phase of the project: the consumers' qualitative remarks were to be used to determine the choice of the most appropriate linguistic tools with which to analyse certain features of the abstracts (to generate hypotheses, in other words); and the consumers' ranks were to be used to provide a rigorous means of comparing and contrasting the results of the analyses (to test hypotheses, in other words).

8.6 Conclusions

This chapter has presented a wealth of data pertaining to the success of abstracts and abstracting. As stated above in section 2.5, different types of data have been collected in a number of different ways to ensure a good level of credibility (one of Lincoln and Guba's canons for successful qualitative enquiry, 1985: 301 - 316), one dimension of which is 'to demonstrate that the inquiry was conducted in such a manner as to ensure that the subject was accurately identified and described' (Marshall and Rossman 1989: 145). This has been effected by triangulating both the data and the way in which it has

been collected, and it is hoped that the subject, having been carefully selected and circumscribed, can now be investigated in subsequent chapters.

Perhaps one of the most interesting results from this phase of the data collection was the finding that neither the lecturer nor the students had much to say of any real linguistic substance concerning the actual language of the abstracts. It is probably not coincidence that realisation seems to be the area where contradictions are most apparent in abstracting standards and guidelines. Part Four below will concentrate on precisely these neglected areas, and will pay close linguistic attention to the actual words the students used.

Several interesting discoveries have been made about the 'abstractability' of the different source texts; a correspondence was evident in the data, for example, between those source texts judged most successful and those most 'abstractable'. Further, insights into the process of abstracting have been documented which are far more detailed and useful than is usual in the normative literature.

The fact that there is so much interesting data, however, is encouraging but potentially injurious. On the one hand, it is encouraging that a variety of credible data has been accumulated, data which is grounded in fieldwork, but which also capitalises on the strengths of other types of enquiry. On the other hand, the sheer bulk of data amassed means that one cannot hope to analyse it all in a similarly trustworthy fashion. As Wolcott (1990: 46) warns in a chapter entitled 'Keeping Going', clearly some focusing of interest is necessary: 'Your major concern ... is to get rid of data - to focus progressively, to home in on your topic. You can note in passing (in literal notes to yourself, or in asides shared with readers) important and intriguing issues you must leave for another time'.

But this focusing must not be arbitrary. The data to be concentrated upon should be that which is most interesting or least well understood. It is submitted that the success judgements fulfil both these criteria.

The abstracts products data in the form of the *ranks* was particularly interesting because it was discovered that informants vary greatly in their preferences. Unfortunately the design of this research does not permit extensive tests of statistical significance, but it seems reasonable to conclude that the success of an abstract is determined not by any one single factor, but by a complex set of factors. Further, it seems reasonable to conclude that informants either do not agree on the composition of these factors, or that they choose to attach differing levels of importance to each.

The abstracts products data in the form of the *qualitative comments* was particularly interesting because it was discovered that the advice offered by the abstracting standards and guidelines literature, even if it were capable of being acted upon, would, at very most, forestall only half of the criticisms levelled by the consumers. READ comments are especially interesting in this regard, not only because they are so widespread (the highest scoring code, 66), but also because they highlight phenomena which receive little or no attention in the literature surveyed in Part Two. It is perhaps not coincidental that the phenomena underlying these types of comment are less well understood linguistically. Subsequent chapters will therefore attempt to redress this imbalance by considering precisely those opinions which seem to matter most to the informants, but which require further linguistic elaboration (see Part Four below).

Part Four

Analysis

Introduction

This chapter marks the boundary between Part Three, which discussed the various means of collecting suitably 'grounded' data (see Chapter 2), and the remainder of the thesis, which constitutes an attempt to make sense of it: to explain how and why some abstracts are considered more successful than others. Although the chapter physically separates the discussion of data collection and data analysis, their interrelationship will be briefly examined, and it is hoped that this chapter will be seen to provide a natural and well-motivated link between the two.

Chapter 2, which argued for the adoption of a multimethod approach, concluded that the design best equipped to answer the most pressing research questions was one which synthesised observational sophistication and theoretical precision. Hopefully observational sophistication will have been evident in Part Three above; these chapters reported on work which was carried out in a predominantly naturalistic style. Theoretical precision requires that attempts are made to falsify explicitly stated hypotheses. But where do these hypotheses come from? And what bearing do they have on the data? What kinds of analysis can be brought to bear upon the research problems?

Given that 'almost every aspect of a survey inquiry can be made the subject of pilot work, ... a line has to be drawn somewhere' (Oppenheim 1966: 25), and to be suitably 'grounded' (Glaser and Strauss 1967), this line should not be drawn arbitrarily, but with a view to explaining the most interesting of the informants' qualitative opinions. Two questions present themselves therefore: which of the many judges' comments are to be concentrated upon, and which particular linguistic tools should be used to explain them?

In answer to the first of these questions, it was thought preferable to focus on precisely those opinions which cannot be sufficiently well explained by

existing knowledge. So, while it would be relatively straightforward to devise abstracting guidelines adherence to which would ward off comments such as:

'Uppercase irritating'

(Judge 4 on Distinct Personality Type, version A);

'No need for citations, I feel'

(Judge 6 on Distinct Personality Type, version B);

'Why contract [years] to "yrs"?'

(Judge 8 on Distinct Personality Type, version C);

'the first sentence has no verb in it'

(Judge 3 on Tanzania, version C);

'too long'

(Judge 2 on Vickery &Vickery Chapter 1, version A);

'Didn't like the headings'

(Judge 6 on Vickery &Vickery Chapter 1, version A),

it would be altogether more difficult to know how best to advise the writers whose abstracts attracted the following criticisms:

'too disjointed'

(Judge 3 on Tanzania, version A);

'incoherent'

(Judge 4 on Tanzania, version A);

'very jerky portrayal of concepts'

(Judge 6 on Tanzania, version A);

'too wordy'

(Judge 1 on Tanzania, version C);

'badly written, unclear'

(Judge 2 on Tanzania, version D);

'I didn't like the style ... too chatty possibly'

(Judge 3 on Tanzania, version D)

'the structure is odd'

(Judge 3 on Vickery &Vickery Chapter 1, version A);

'is very difficult to read'

(Judge 1 on Vickery &Vickery Chapter 2, version C).

In answer to the second of the above questions, how best to investigate these comments, a short-list of types of linguistic analysis had to be constructed,

from which selections could be made.

In Part Four, a systemic-functional grammar (SFG) perspective has been adopted because, as even the sternest critics of Halliday 1985 acknowledge (see for example Hudson 1986: 801 and Huddleston 1988: 137 - 140), SFG is particularly well suited to the analysis of *naturally occurring texts*. This was considered to be a major advantage of the theory. Which types of analysis should be considered so as to shed most light on abstracts and abstracting? Halliday 1985 is a convenient starting point for a short-list of potentially useful types of analysis:

- 1) Theme (1985: Chapter 3);
- 2) Mood (1985: Chapter 4);
- 3) Transitivity (1985: Chapter 5);
- 4) Groups and Phrases analysis (1985: Chapter 6);
- 5) Clause Complex analysis (1985: Chapter 7);
- 6) Intonation and Rhythm (1985: Chapter 8);
- 7) Cohesion (1985: Chapter 9);
- 8) Grammatical Metaphor (1985: Chapter 10).

Further techniques which other researchers have found helpful, but which are not treated in Halliday 1985, include analyses of:

- 9) Discourse structuring (according to Rhetorical Structure Theory, or RST, for a convenient summary see chapter 4 of Fox 1987; and / or Generic Structure Potential, or GSP, see for example Hasan 1985/1989a¹; and / or the type of Genre Analysis advocated by Swales 1990);
- 10) Metadiscourse (see for example Crismore 1984, Vande Kopple 1985, and Stainton, forthcoming (b));
- 11) Thematic Progression, as evinced by the Prague School, and
- 12) Lexical Density (see for example Ure 1971, Halliday 1989).

Of these, 6) can be discounted forthwith, as it primarily concerns spoken discourse. In addition, it is not immediately obvious how grammatical metaphor, though doubtless a useful tool for the investigation of both modes of language (see for example Jones 1988, who examines academic writing), can

¹This article, originally written in 1980, appeared in a book published by the Deakin Press in 1985, and was subsequently reissued by the Oxford University Press in 1989.

help shed light on, or, in anthropological terms, 'sensitize', the particular qualitative remarks presented above. Mood would not be expected to reveal too much of interest either, since it is unlikely that the different abstracts could be differentiated by those components of meaning concerned with the exchange between speaker and hearer, the archetypal speech function of the abstract being to furnish information. This is not to say of course that these aspects of language have *no* relevance for the problem in hand. However it is beyond the scope of this thesis to investigate every feature of language which may influence the perceived success of an abstract. Clearly some process of homing in on particularly interesting areas is required.

One particularly interesting finding from Chapter 8 of the present thesis was that many of the judges' complaints have to do with the notion of coherence. Hasan has argued (Hasan 1985/1989a: 52) that, to be coherent, texts must exhibit both unity of texture and unity of structure.

Unity of texture can be investigated by making appeal to thematic structure and cohesion (Halliday 1985: 313). Unity of structure can be investigated by making appeal to the types of analysis outlined in 9) above.

When the judges flag abstracts as being inappropriate stylistically, typically they object to 'chatty', or 'verbose' language. The lexicalisation 'chatty' suggests that the abstractors may be using features of language more appropriate to speech than writing. Halliday 1989 has compared and contrasted speech and writing, arguing that both can be considered complex along separate, though related, axes. The first of these dimensions, the notion of lexical density, refers to the relative proportion of lexical and grammatical words in a text. The second of these dimensions, the notion of grammatical intricacy, refers to the degree to which grammatical units combine univariately to form larger grammatical units. These axes can provide measures of the lexical and grammatical complexity of a text, and so it is hoped that some of the complaints relating to 'chattiness', 'verbosity' and the like can therefore be accounted for.

One of the judges asked to evaluate the General Knowledge abstracts mentions that one abstract 'uses words like "predicting", "suggests", "considers", "reveals" and "finds" to describe the contents'. These verbs realise particular kinds of *process* (Halliday 1985: 101 - 102), and so a study of transitivity might help to explain this informant's intuitions. And, since another states 'too many phrases like "the chapter then sets the tone for ... " etc - superfluous', an analysis of metadiscourse might be in order.

Unfortunately, time and space limitations dictate that not all of these areas

can be given the attention they deserve. Thematic Progression will therefore not be considered here, as Stainton (forthcoming (c)), for example, has found it to be less useful than the Hallidayan treatment of theme in differentiating between successful and unsuccessful texts. Also, the metadiscoursal categories proposed are currently less than 'precisely relatable to their exponents in the data' (Sinclair and Coulthard's second criterion for descriptive systems, 1975: 16), so this too will not be taken further.

The short-list has now been whittled down to: generic structure, Theme (Hallidayan approach), transitivity, grammatical intricacy, lexical density and cohesion:

- Generic structure will be investigated below in Chapter 9 for two reasons: firstly, coherence appears to be a strong discriminator between judges' opinions, and, as has been mentioned above, Hasan has motivated generic structure as one of three different types of analysis which can be used to investigate what she refers to as the two unities - unity of texture (as measured by analyses of Theme and Cohesion) and unity of structure - which are necessary preconditions for coherent text; secondly, Swales 1990, using a somewhat different approach to genre, has discussed abstracts explicitly, albeit in a rather programmatic way. In fact, this second type of approach will be explored in Chapter 9;
- Lexical density will be investigated in Chapter 10 since, not only does it appear useful for assessing lexical complexity (and with only minimal changes can be used to track lexical variation), it also forms the first step of the next type of analysis;
- This next type concerns cohesion and is investigated in Chapter 11. Although there are many references both in the linguistic and education literature to suggest that analysis according to Halliday and Hasan 1976 is unworkable in practice, a recent monograph (1990) by Parsons, again addressing questions of success, reports that Hasan's later work (1984) on cohesive harmony has proved valuable;
- Clause complex analysis will be carried out in Chapter 12, since, not only does it appear useful for assessing grammatical intricacy, it is a necessary preliminary to the next type of analysis;

- An analysis of Theme is carried out in Chapter 13. This largely follows Chapter 3 of Halliday 1985, with some modifications.

Transitivity will not be taken further. Although it has been shown to pay dividend (see for example Martin and Peters 1985: 69 - 72, or the seminal treatment of William Golding's 'The Inheritors' in Halliday 1973), the decision was taken not to pursue transitivity primarily because there were *comparatively few* qualitative remarks which motivate its further investigation.

The paragraphs above have discussed why certain descriptive frameworks were chosen. The different kinds of analysis will be investigated in the five chapters which follow. Before these analyses are embarked upon, however, it is proper to make perfectly clear how they are to be viewed, and what they are intended to accomplish. The remaining paragraphs in this introduction will briefly review how these analyses were attempted, and will give an indication of their relative scope.

Two major problems have shaped the presentation of the work in the following chapters.

The first major problem was that *some of the descriptive frameworks were found to be more useful and usable than some of the others*. Consequently more time and effort was spent investigating those analyses which were thought to hold the greatest explanatory promise.

The analyses described in Chapters 9 (Genre) and 11 (Cohesion) were less successful. In these two chapters, only two abstract versions are analysed. Further analysis was abandoned once it had become clear that serious problems concerning the descriptive frameworks meant that not all of the data could be analysed satisfactorily. Chapters 10 (on Lexical Texture), 12 (on Grammatical Intricacy) and 13 (on Theme), on the other hand, were less problematic, and so the analyses were taken further. All seventeen of the Information Science abstracts were analysed for each of these three types of analysis.

The second major problem was that *the effects of the different linguistic phenomena investigated are not equally well understood*. In some cases, similar hypotheses have already been investigated; in other cases, there is little pre-existing knowledge available.

In cases where there is an established body of knowledge concerning a

particular kind of analysis, it was thought more appropriate to follow a more deductive style of analysis (on the difference between deduction and induction, see section 1.4 above), stating hypotheses at the outset of the investigation and then subjecting these to rigorous attempts at falsification. This has been the predominant pattern for Chapter 10 (on Lexical Texture) and for Chapter 13 (on Theme).

However, in cases where little was known concerning the effects of particular linguistic variables, it was thought more appropriate to follow a more inductive style of analysis, one which 'allow[s] the important analysis dimensions to emerge from patterns found in the cases under study without presupposing in advance what the important dimensions will be' (Patton 1990: 44). This has been the pattern for the later sections in Chapter 12 (on Grammatical Intricacy). Such a combination of inductive and deductive strategies is an essential part of analysis (Strauss and Corbin 1990: 148), and should lead to more trustworthy findings.

Having said that three of the five chapters in Part Four are on a larger scale, it must still be remembered that this whole thesis is meant as a kind of 'macro pilot analysis'. Because the relevant dimensions of perceived success have yet to emerge for abstracts and abstracting, it is appropriate that a number of different types of analysis are investigated in breadth. This thesis is meant to identify promising avenues for further investigation; once the relevant determinants of writing quality have been identified, it would then be appropriate to investigate how exactly these affect readers' perceptions in a more focused way, controlling variables whenever possible. However, this type of deductive enquiry is beyond the scope of the current work. The primary purpose of this thesis is to *enable* more directed research, by providing a well-motivated means of narrowing down the search for the relevant underlying drivers of readers' preferences.

9

Analysis of Generic Structure

9.1 Introduction

This chapter investigates the degree to which generic structure predicts the informants' preferences; section 9.2 discusses why generic structure was thought to be a particularly appropriate type of analysis. A brief overview of the theory underlying genre is given in section 9.3: section 9.3.1 provides a discussion of the general issues in genre study; and, in 9.3.2, some published observations concerning abstracts in particular are evaluated.

The pilot analysis is discussed in section 9.4 onwards, starting with the reasons why two abstracts in particular were selected for analysis. One possible interpretation for Distinct Personality Type, Version D is given in section 9.5, and one for Distinct Personality Type, Version C in 9.6. There is a brief discussion of the analyses in section 9.7, prior to the conclusions in 9.8.

Because so many analytical problems were encountered in the pilot analysis of generic structure, it is proper to preface the following discussion with a short statement pointing out what readers can and cannot expect to gain from reading further.

The following analysis highlights differences in the structure of the two abstracts and in the way the structure is signalled. These differences help to explain some of the judges' qualitative remarks. Also, two particular variants of generic structure (proposed by Graetz 1985 and by Hill et al 1982; see below) were found to be accurate in predicting which elements of generic structure have to be present in order for an abstract to be perceived to be successful. Further, there is evidence to show that the ordering of the elements can affect perceived quality.

However, it was concluded that this type of analysis could not be taken further, principally because, currently, the framework cannot always be applied with the degree of reliability required for precise hypothesis testing.

9.2 Why Analyse Generic Structure?

There are two reasons why it was considered potentially useful to analyse the generic structure of the abstracts.

Firstly, informants seem to be particularly sensitive to structural issues: when the qualitative opinions were analysed, 'STRUC' emerged as the fourth highest scoring code, with a total of 35 (see section 8.5.3.2 above). In many cases, informants mentioned the word 'structure' explicitly. A small selection of these comments appears below:

concerning Tanzania, version D:

'I find E better structured and easier to read' (judge 6);

concerning Vickery and Vickery, Chapter 1, version A:

'is totally hopeless - the layout is confusing and has no real structure' (judge 1).

Also, some informants refer to elements of structure explicitly:

concerning Tanzania, version A:

'well constructed overall, with a brief introduction and conclusion' (judge 8);

concerning Tanzania, version E:

'E is informative, easy to read and has a beginning, middle and end' (judge 3);

concerning Distinct Personality Type, version E:

'would have benefitted from better conclusion (the author's conclusion is given in the third sentence and the remaining half merely elaborates on this' (judge 8).

Secondly, existing research carried out under more controlled circumstances has shown that structure affects readers' perceptions of the quality of summaries. In an interesting article, Stanley (1984) adopts Hoey's problem-solution model of discourse structure (most comprehensively articulated in Hoey 1983), and shows how, when informants were asked to rank the competing summaries in terms of 'quality', structurally 'complete'

summaries were almost always preferred over those missing either problem or solution sections. Although there are a number of methodological differences between Stanley's work and this study - Stanley's informants were asked to read the source material as well as the summaries, and the summaries were not naturally occurring, but were written by the analyst herself - there is good reason to believe that the absence of certain structural elements engenders feelings of dissatisfaction or disorientation in the reader.

Having motivated generic structure as one possible means of reconciling the abstracts with the success judgements, a brief review of genre study will now be provided.

9.3 Problems with 'Genre'

The word 'genre' means different things to different people. The concept is discussed in many different academic fields and at varying levels of specificity. Definitions abound in the literature but it is fair to say that there does not yet exist a wholly satisfactory theory of genre. As this is a pilot study, it is beyond the scope of this chapter to provide a critique of the different theories, tools and techniques used to study different types of writing. The reader is referred to Swales 1990, a book-length treatment of genre study, containing useful reviews of work in different fields, together with many references to more specifically linguistic kinds of work.

The title of section 9.3.1 below has been deliberately chosen to acknowledge its reliance on another useful account, Stainton (forthcoming (a)). Section 9.3.2 concerns abstracts in particular and considers a number of observations made by Swales (1990) concerning what are claimed to be their distinguishing characteristics.

9.3.1 A Summary Overview of Genre and Genre Study

Genre is a thorny concept to have to grapple with, partly because there are fundamental differences of opinion even between people with similar academic persuasions. Within systemic linguistics, for example, Butler (1985a: 88) points out that Gregory (1967) wishes to account for genre primarily in terms of tenor, Halliday (1978) primarily in terms of mode, and Berry (1982) primarily

in terms of field. Similarly, some workers equate genre with register (Hasan 1978), while others insist on their separation (a position held by Martin in various publications; see for example Martin 1984, 1985, and Martin and Rothery 1986). Ventola 1984 provides useful commentary on this disparity, while Couture 1986 offers the beginnings of a solution, arguing that registers determine lexical/grammatical choices in texts, while genres determine discourse choices.

Stainton (forthcoming (a)) examines the 'plethora of definitions' and suggests that each may be usefully placed on a contextual continuum: at one end would lie those theories which exclusively concern the features of the texts under scrutiny; at the other those which exclusively concern the contexts in which the texts are situated. This continuum can be diagrammed as follows:

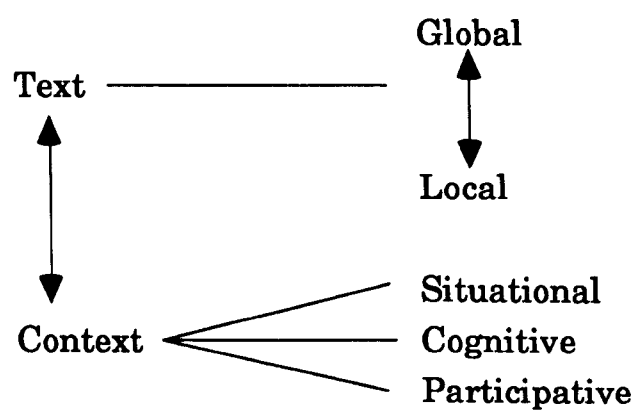


Figure 9.1: The Contextual Continuum
(after Stainton forthcoming (a))

In practice most accounts fall somewhere between the two extremes. Stainton cites the following studies among others as combining aspects of each: Fries 1987, Olson et al 1981, and Mosenthal 1985. Within textlinguistics, perhaps the greatest strides towards bridging the gap between text and context have been made by Hasan in her work on Generic Structure Potential (GSP), which is 'a realisation of the values of the contextual configuration, a specific realisation of field, tenor and mode values of the discourse' (Stainton forthcoming (a)).

Hill et al describe a study designed to help ESL students read and write experimental research papers. Their basic premise is that, by paying attention to what they call rhetorical divisions, students will become sensitised to organisational features, and so will be more successful in reading and writing such texts.

Hill et al demonstrate how a text can be partitioned rhetorically with reference to an article from psychology. They motivate this choice 'because it represents a standard rhetorical model in the sciences, but papers in other fields could easily be used in the EFL classroom as well' (1982: 335). In many ways, the model they adopt is similar to that discussed by Swales for the Research Article (RA), although there are one or two minor differences. Theirs is basically a tripartite model - Introduction, Procedure, Discussion - with Procedure being further subdivided into Methods and Results. In addition, they say that psychology papers can be conceived of in terms of an hourglass shape:

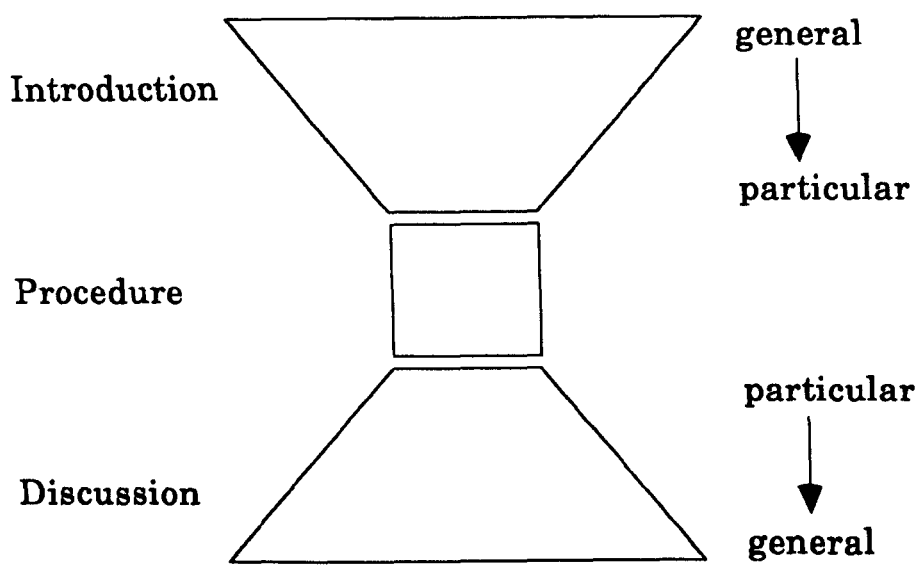


Figure 9.2: The Generic Structure of the Psychology Paper
(from Hill et al 1982: 335)

This diagram is meant to convey how the article progresses firstly from the general to the particular, and secondly from the particular to the general. This hourglass shape, they argue, is a direct consequence of the functions of the different generic elements: 'The introduction provides a transition from the larger academic field to the particular experiment. The procedure section describes the particular experiment. The discussion, like the introduction, is a transition too, but its purpose is to guide the reader from the particular experiment back to the larger academic area' (1982: 334).

Most linguistic studies concern either the organisation of text at a global level or at a local level. Hill et al's concerns are with the global. Hasan's work

on GSP is a major contribution to the understanding of global discourse structuring, but falls short of demonstrating exactly how these structures determine more local choices. Biber's work (1986, 1988), on the other hand, provides a useful account of how certain local, distinctive features cluster together in particular types of language, but does not relate these to higher constraints. It is the reconciliation of these two levels which will lead to greatest progress in understanding genre.

The global/local continuum is also relevant to the work presented in this thesis. The pilot analysis presented in this chapter attempts to provide a characterisation of the global organisation of two of the abstracts. The remaining chapters in Part Four, however, discuss analyses at altogether more local levels: Chapter 10 discusses lexical texture; Chapter 11 discusses cohesion; Chapter 12 discusses grammatical intricacy; and Chapter 13 discusses Theme.

9.3.2 Abstracts in Particular

This section discusses Swales' account of abstracts (1990: 179 - 182). He makes the following introductory observation (1990: 179): 'the essence of the genre is one of distillation. Essentially, it is this distilled quality that gives abstracts their particular character and makes them easy to recognize'.

This distillation, he claims, determines the nature of the abstract at both local and global levels. Turning to the local level first of all, Swales cites Graetz (1985: 125) who claims that the abstract is characterised by:

<u>Use of:</u>	<u>Avoidance of:</u>
Past Tense	Subordinate Clauses
3rd Person	Abbreviations
Passives	Jargon
Non-negatives	Symbols
	Superlatives
	Adjectives
	Illustrations
	Footnotes, etc.

Swales comments 'Although Graetz's corpus consisted of 87 abstracts drawn from a variety of fields, her use of the generic definite article [*the* abstract is

characterised by ...'] is perhaps a little bold' (1990: 180), and goes on to show where he agrees with her claims and where he disagrees.

The corpus of abstracts examined in the present study consists of only seventeen Information Science abstracts, and yet it is possible to provide counter-examples to virtually all of Graetz's claims. For example, one of the abstracts considered below (Distinct Personality Type abstract D, the version which was considered the most successful of the set) relies more on the present tense than any other; actives are more common than passives in many of the abstracts; and statements involving "not"s, for example, regularly occur. Perhaps the most easily falsifiable hypothesis is that writers avoid using subordinate clauses in abstracts. One of the findings from Chapter 12 below, concerning grammatical intricacy, is not only that the abstract writers studied in this research make heavy use of subordination, but also that readers seem to prefer those abstracts with higher levels of grammatical intricacy. Graetz's observation that third person is used in abstracts is supported by the Information Science data, but that is hardly unexpected.

At a more global level, Graetz postulates the following structural elements for the abstract: Problem - Methods - Results - Conclusions. Swales disputes this, however, arguing that the structure for the Research Article (RA) - Introduction - Methods - Results - Discussion - applies equally well to the abstract (1990: 181): 'Although further research is needed, it seems to be the case that most abstracts reflect the IMRD pattern of the RA itself, allotting a sentence or two for each section'.

This claim will now be investigated with reference to two Information Science abstracts.

9.4 Selection of Texts for Pilot Analysis

This section discusses why abstract versions C and D of the Distinct Personality Type original were selected for analysis.

The Distinct Personality Type set was chosen after a reading of Hill et al 1982. This was because the Distinct Personality Type original text is a review of psychometric theory, and so a useful starting point for the analysis is to assume that it will conform to the hourglass shape diagrammed in Figure 9.2 above. The Distinct Personality Type set will therefore be investigated to see to what extent the abstracts reflect the underlying structure, which, it is hypothesised, successfully characterises the original text.

Distinct Personality Type abstract versions C and D have been selected for pilot analysis because they have attracted more 'STRUC' criticisms than the other versions; these comments will be listed below at the beginning of the following two sections. In addition, D is the most preferred overall, while C is the next-to-least preferred, so this particular pairing represents a useful spread of perceived quality. (One of these abstracts, version D, will also be used in the cohesion pilot analysis described below in Chapter 11).

9.5 One Possible Interpretation of Distinct Personality Type, Abstract Version D

This section presents one possible analysis of the generic structure of Distinct Personality Type, abstract version D. There are two points to be noted with regard to the following discussion.

Firstly, the analysis is altogether more intuitive, or subjective, than those presented in the following chapters in Part Four. This is because, at present, the formal mechanisms underlying the realisation of generic structure are not well understood, and it is more difficult to guarantee analytical reliability and validity the further one moves away from local to more global features of text organisation.

Secondly, discussion will proceed sentence by sentence, the assumption being that, when reading texts, people continually reconcile their perceptions of how the text is (and has been) organised (in previous sentences) with their expectations of how the text will be organised in future sentences. However, no great claims about psychological units of processing should be inferred from what has just been said. The orthographic sentence is used merely as a convenient presentational device.

Turning now to the judges' qualitative opinions concerning Distinct Personality Type, abstract D, those coded 'STRUC' (glossed as 'The arrangement and interrelationship of parts (or the lack of them) in the abstract. Construction. Organisation of the text, for example, Introduction, Body and Conclusion. Ordering of these elements'; see Chapter 8 above) are as follows:

'wins on length, wording is concise yet informative, ends with conclusions'
(judge 1);

'seemed to extract key points and present them in clear ordered way' (judge 2);

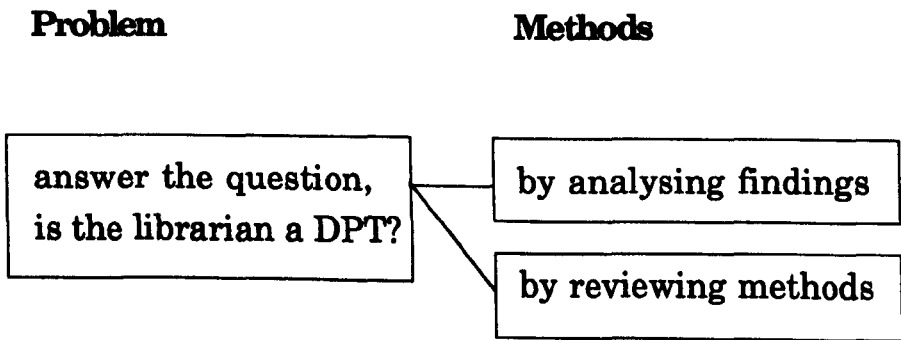
'Concise introduction, middle and conclusion' (judge 8).

In what follows, sentences from abstract D will be presented one by one, each accompanied by a statement of the analyst's intuitions, the idea being to see to what extent the analysis can explain the qualitative remarks listed above.

Sentence 1

The article sets out to answer the question, is the librarian a distinct personality type?, by a process of analysing the findings of psychological research conducted over the last 30 years, and a critical review of the methodologies used.

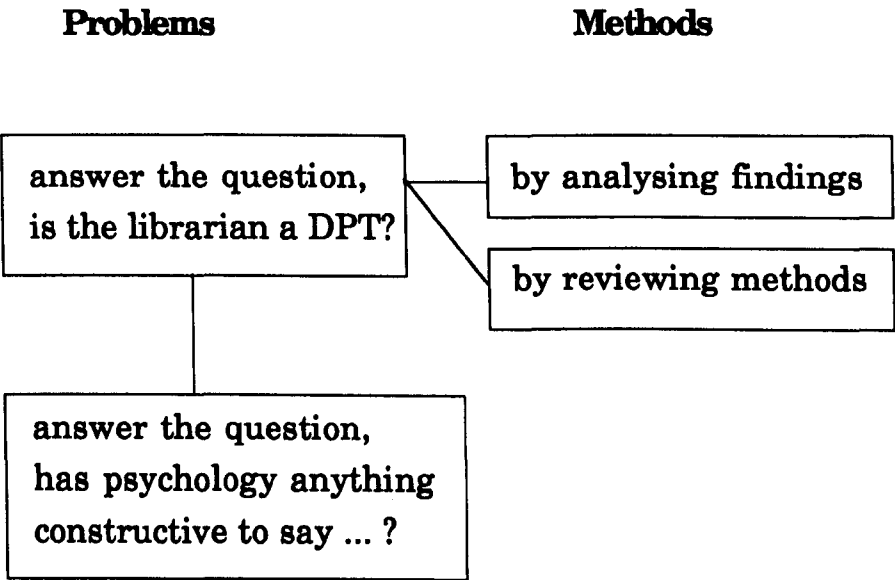
As stated above in Chapter 5, the abstracting authorities place great store in the topical sentence. Version D's first sentence frames the entire abstract by providing an indication of the intent of the original article; it constitutes a kind of meta-abstract, in other words. Sentence 1 is to be regarded as essentially a statement of a problem, the problem being to answer the fundamental question, 'Is the Librarian a Distinct Personality Type?', a question which the original author has chosen to be the title of his article. Sentence 1 also contains an adverbial phrase with a rankshifted clause inside it. The 'by a process of ... ' formulation lexically signals that the following information will detail the methods by which the question is to be answered. There are therefore two different types of information realised in the first sentence, which will be diagrammed as follows:



Sentence 2

The author also addresses the question whether or not psychology has anything constructive to say about the library as an organization.

The grammar of this sentence is remarkably similar to that of the previous sentence: both can be said to have a non-topical Theme (see Chapter 13 below); both contain mental processes; and in both cases the yes/no questions are stated explicitly, using an appositive relation. The 'also' is perhaps a signal that no new elements of generic structure are being introduced; what is being discussed is another facet of the problem situation. Sentence 2 is different in that the methods by which the second question is to be answered are not made explicit:



So far, then, the abstract seems to uphold some of the claims made by the different genre theorists: Graetz's first two elements of structure are 'Problem' and 'Methods'; and Hill et al's are 'Introduction' and 'Procedure', the former serving to highlight problems with previously held theories (1982: 335), the latter being further subdivided into 'Method' and 'Results' (1982: 336 - 338).

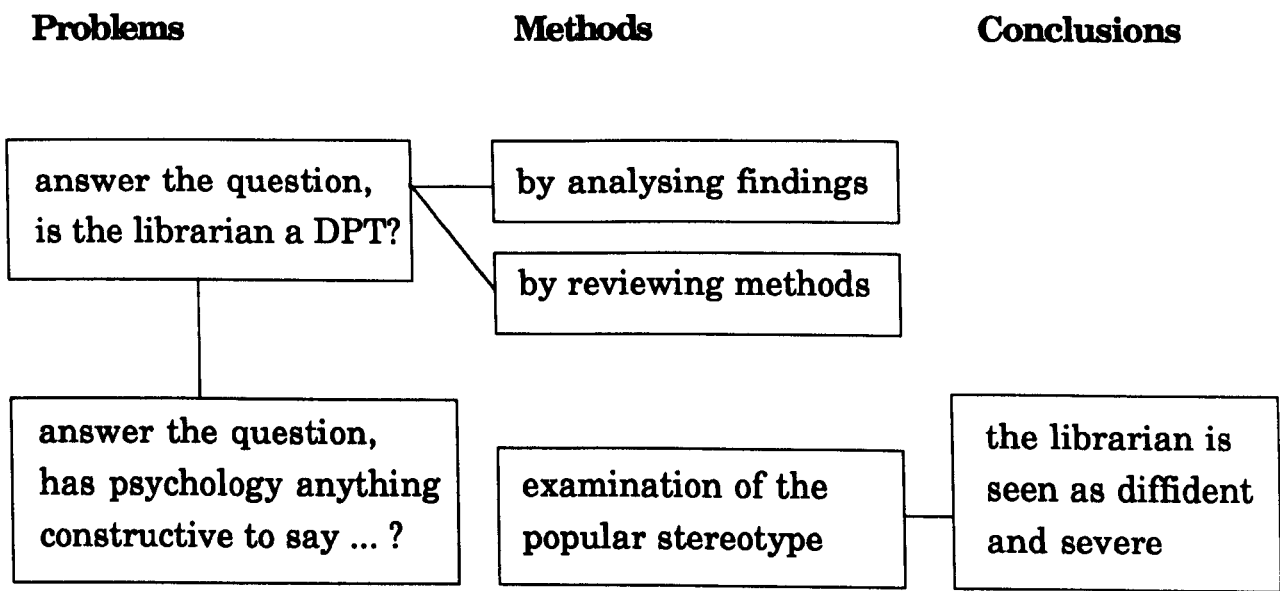
Sentence 3

The popular stereotype is examined and the author concludes that, in popular imagination, the librarian is seen as both diffident and severe.

According to the intuitions of the analyst, sentence 3 is more difficult to interpret, perhaps because it is not immediately obvious how it relates to what

has gone before. The abstractor has chosen to insert a paragraph boundary between sentences 2 and 3, which might prompt the reader to expect a new element of structure. However, it is difficult to accommodate this new information within the existing generic structure. This may explain the comments of the fourth judge, who said 'not enough information. What is the "popular stereotype"?', although, according to the analyst's intuitions, "popular stereotype" is readily interpretable; it refers cataphorically to information realised in the following projected clause.

One possible interpretation for sentence 3 (motivated quite simply by the main verb 'concludes') is that it represents a third element of structure - conclusion - together with an indication of the methods by which the conclusion was reached (motivated by the main verb 'examined'):



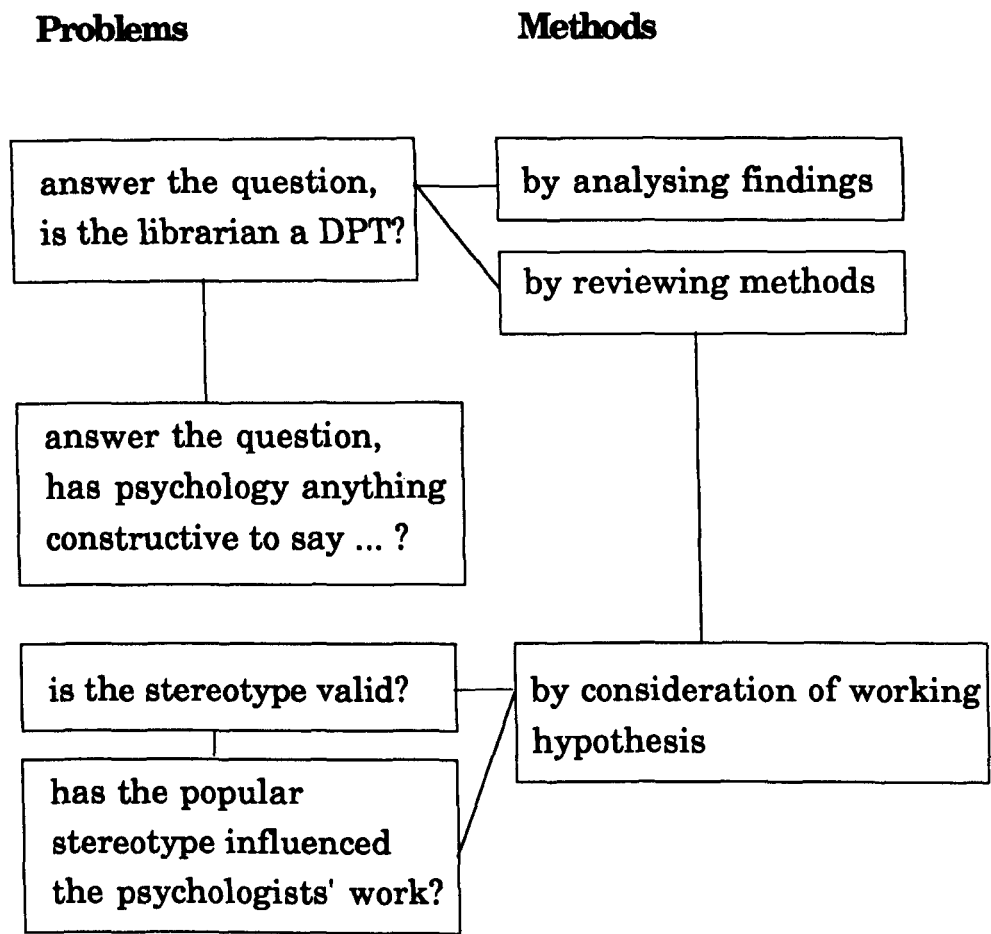
No lines are drawn between the last two boxes and the first four to signal the analyst's intuition that there is something of a break in the chain of argument. In any event, this interpretation will be revised below.

Sentence 4

Using this hypothesis David Fisher reviews the major studies of the last three decades to ascertain how much fact there is in the stereotype, and how far the researchers have been affected by stereotypes.

The first dependent clause of this sentence forces a reconsideration of the structure drawn so far. What was previously taken to be a conclusion is now to be interpreted merely as a working hypothesis, investigation of which will help

to answer two further questions, i) does the (popular) stereotype really exist?, and ii) if so, have the psychologists themselves been affected by it?.

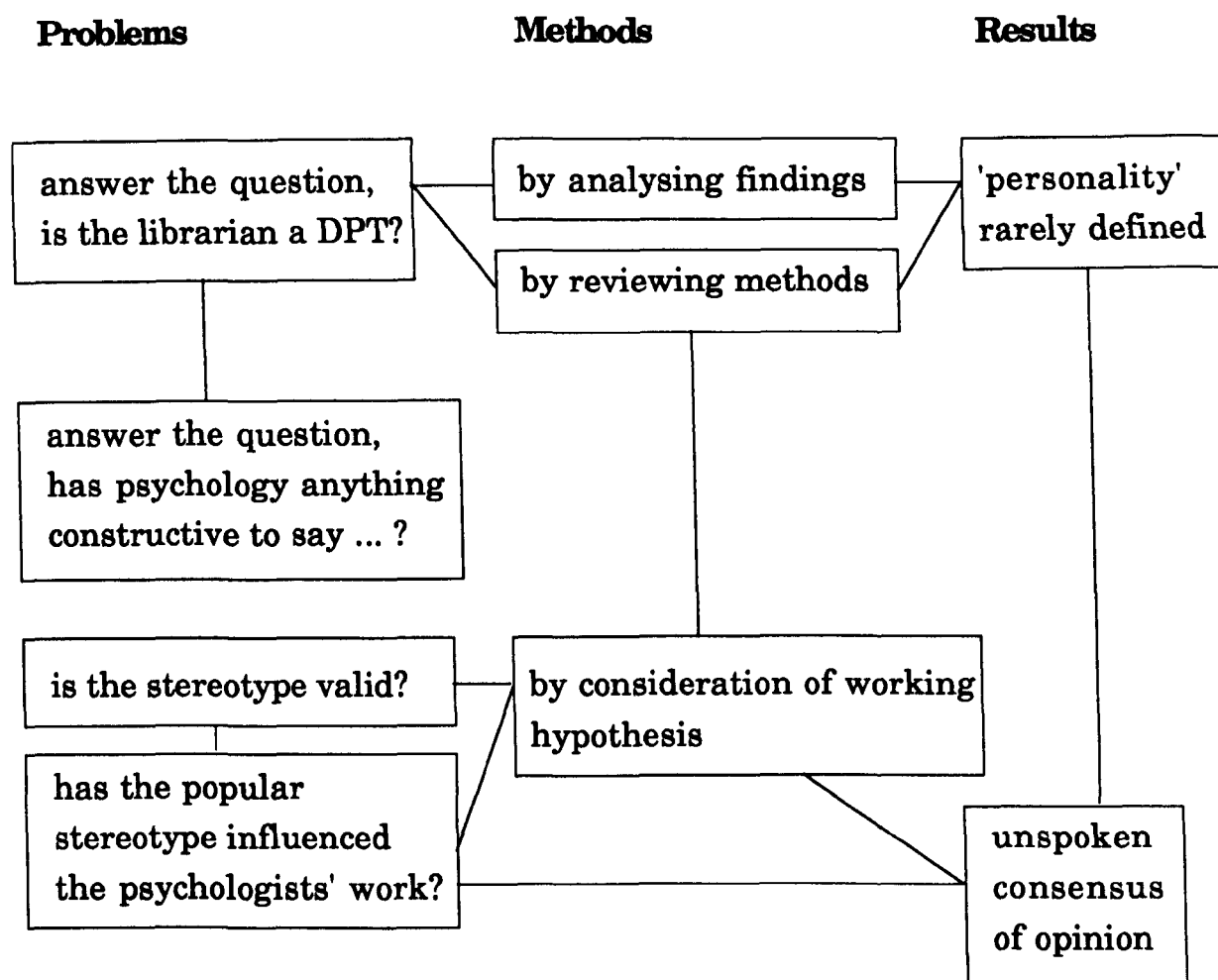


So far, then, there appear to be two blocks of problems, the first pertaining to psychological theory, the second to popular opinion. The link between the two is via the methods: reviewing the literature is intended to solve both sets of problems.

Sentence 5

In discussing personality the author points out that the word is rarely defined, and that most of the researchers assume an "unspoken" consensus of opinion of meaning.

In sentence 5, the two projected clauses (see Chapter 12 below) can each be thought of as being 'results':



These two results sections further help to tie together the two argument blocks. One possible interpretation is as follows: there is insufficient objectivity in the psychological theory (block 1); the psychologists are themselves subject to the popular myths about librarians; and they have let the consensus of (popular) opinion prejudice the nature of their investigations (block 2).

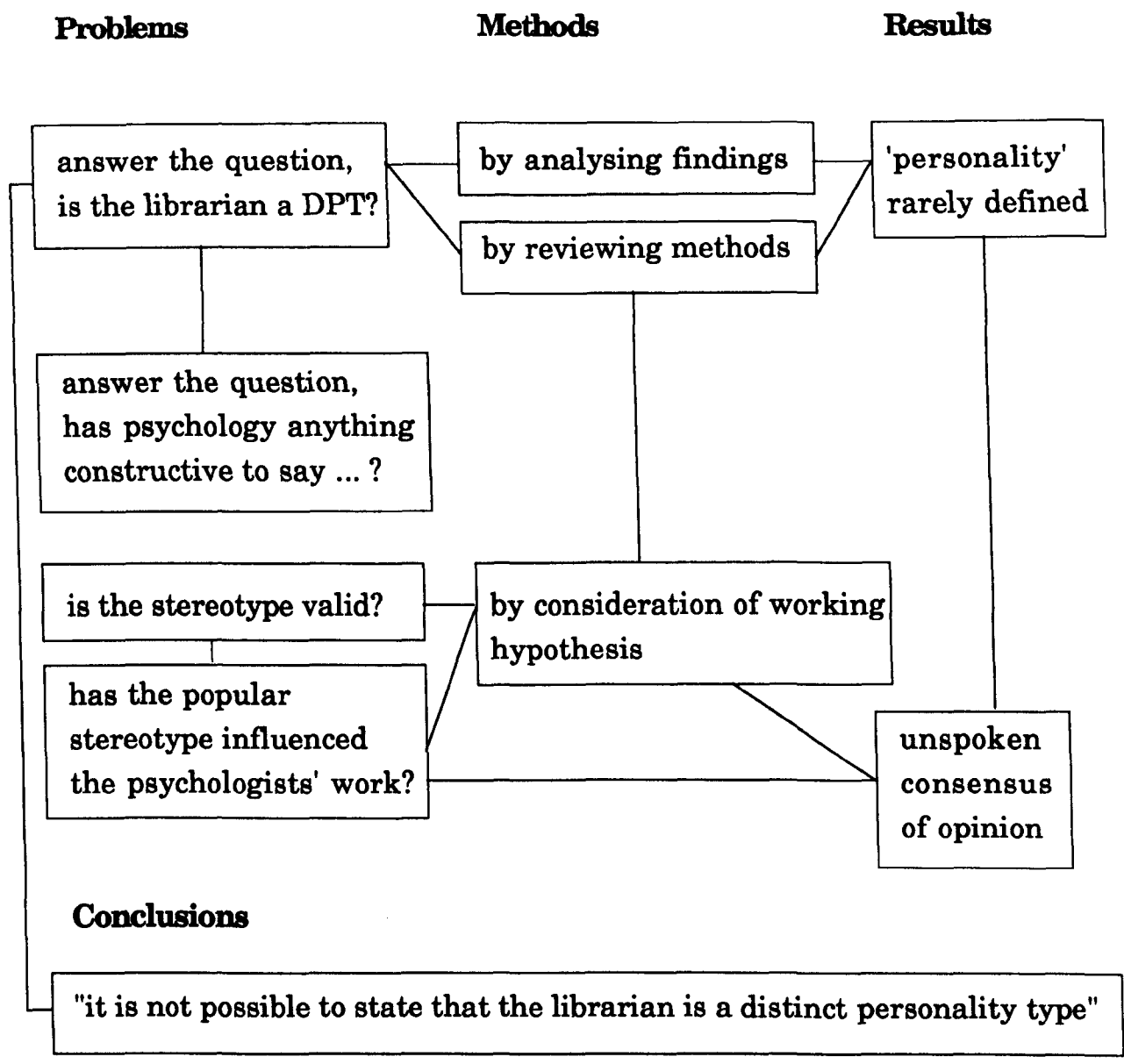
Sentence 6

David Fisher is highly critical of the research methods used, and concludes that on the evidence reviewed "it is not possible to state that the librarian is a distinct personality type".

Sentence 6 starts the third and final paragraph of the abstract. Once again, the new paragraph might set up an expectation of a new element of structure; in fact the bulk of this last paragraph is taken up with concluding material.

The paragraph starts with an evaluative summary of the original author's attitude to the methods employed, and one of the conclusions from the original is quoted directly. This quotation directly addresses the first problem, as set out

in the first sentence:

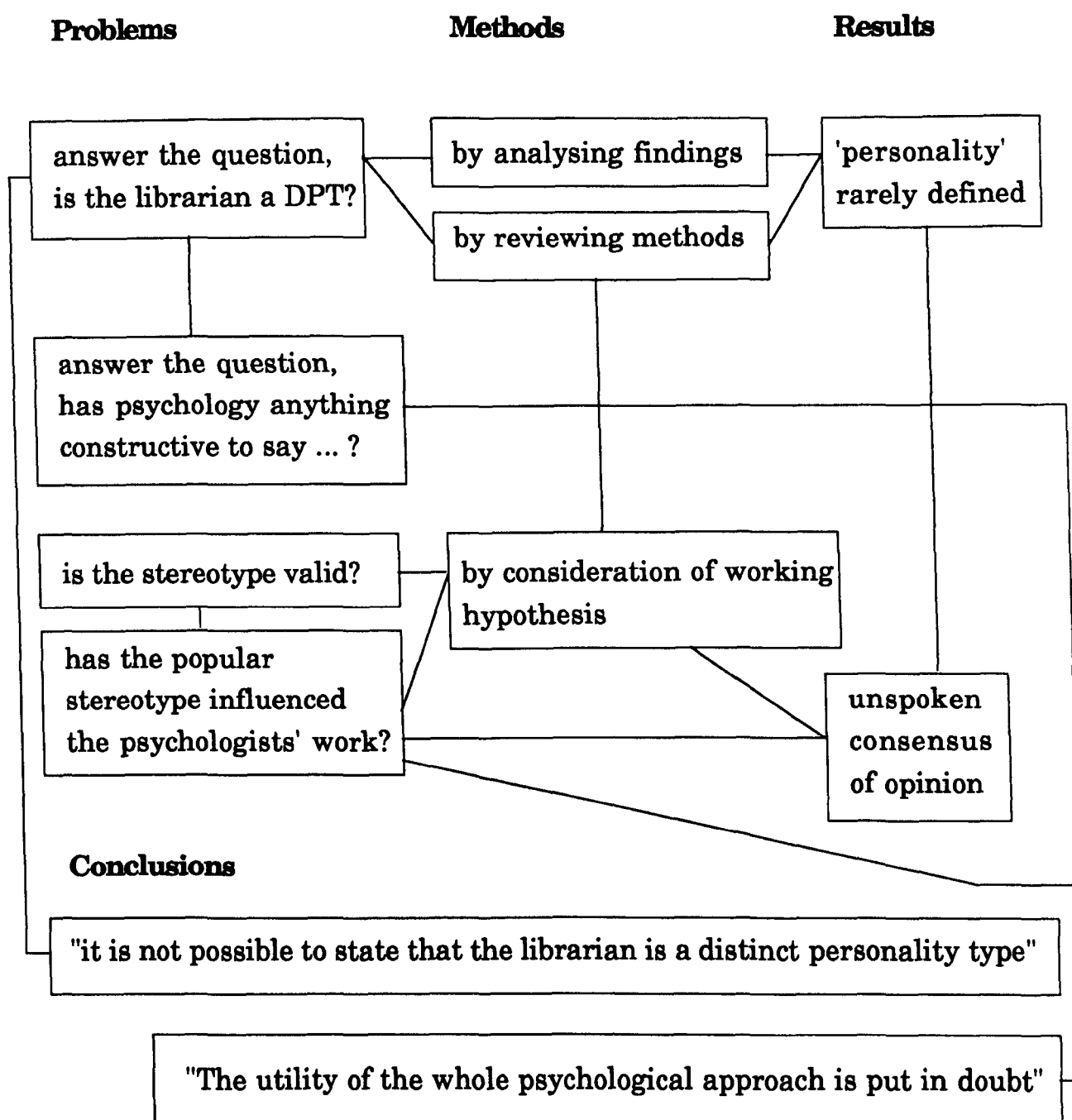


Sentence 7

"The utility of the whole psychological approach is put in doubt".

Once again, this material has been directly imported into the abstract from the original, most probably from its concluding sections.

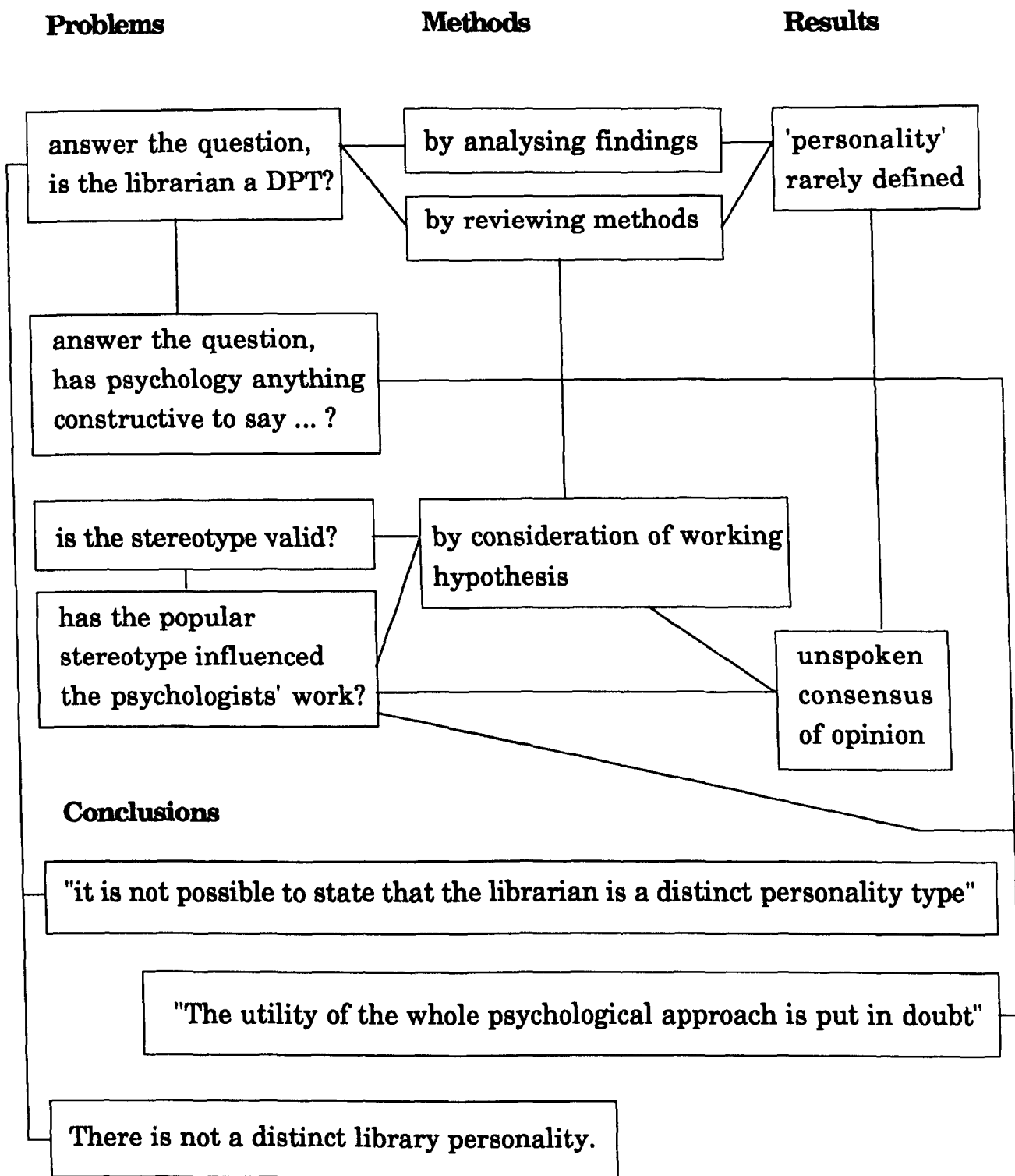
This second conclusion is more general than the first. It seems likely that this conclusion is an indirect answer to the second of the problem questions stated in the introduction, and possibly the fourth as well:



Sentence 8

There is not a distinct library personality.

This final sentence amplifies the conclusion reached by sentence 6, but more completely rounds off the text, the abstractor having chosen to present the author's viewpoint as established fact 'invisibly' using an empty 'there' subject (Davies 1988: 15). This conclusion provides the solution to the most basic problem of the original, by answering the question of the title directly. The final schematic diagram for the generic structure of version D is therefore as follows:



Turning once again to the judges' qualitative opinions, it seems that they have responded well to the abstractor's chosen method of structuring her text. Judge 1, for example, recognises and applauds the multisentential concluding section ('ends with conclusions'). Further, the fact that the analyst had trouble with only one sentence (sentence 3) may explain the perception that the abstract was clearly ordered ('seemed to extract key points and present them in clear ordered way'). All other sentences contribute well to the unfolding

argument. This is in contrast to abstract version C, which will be discussed presently.

The last diagram shows that Distinct Personality Type abstract D can be adequately described by two of the proposals for generic structure outlined above: Distinct Personality Type abstract D, which is the most successful abstract of the set, seems to conform best to the structures described by Graetz: Problem - Methods - Results - Conclusions; and by Hill et al: Introduction (serving to point out the problems with previously held theory) - Procedure (consisting of Methods and Results) - Conclusions.

Swales' account: Introduction - Methods - Results - Discussion (1990); and Hoey's problem-solution framework: Situation - Problem - Response - Result - Evaluation (see for example Hoey 1983: 82) seem to be less useful for characterising the perceived differences in success between different competing abstracts.

These various proposals will now be evaluated further with respect to Distinct Personality Type, abstract C.

9.6 One Possible Interpretation of Distinct Personality Type, Abstract Version C

The judges' qualitative opinions concerning Distinct Personality Type, abstract C which are coded 'STRUC' (glossed as 'The arrangement and interrelationship of parts (or the lack of them) in the abstract. Construction. Organisation of the text, for example, Introduction, Body and Conclusion. Ordering of these elements'; see Chapter 8 above) are as follows:

'loses out because of poor conclusion' (judge 1);

'better ordered but still not clear as to salient points of article' (judge 2);

'No clear plan/progression evident' (judge 8).

In what follows, sentences from abstract C will once again be presented together with statements of the analyst's intuitions, the idea being to see to what extent the analysis can explain the qualitative remarks listed above.

It is altogether harder to infer any kind of structure in this abstract; this may in part explain its relatively low preference marking. However, an

attempt will be made to apply the various generic structure proposals to the abstract with a view to discovering what has gone wrong, as it were.

Sentence 1

Psychology was first used to study the personality characteristics of librarians and the effect these had on their library service and organisation by Alice Bryan.

This first sentence cannot be considered to be a meta-abstract, but rather situates the theory historically and simultaneously introduces one of the main protagonists. Swales regards the historical introduction as one means of beginning a Research Article, but in contrast to Hill et al and Graetz, this first sentence contains nothing that can be considered to be a problem. This differs markedly from abstract D, in which a set of problems were explicitly introduced from the very beginning.

Sentence 2

She suggested a distinct personality type.

It is difficult to know which element of generic structure is realised by this sentence. The mental process verb "suggested" is characteristic of what Labov (1972a: 372 - 373) terms 'embedded evaluation', but it does not answer 'how successful was this?', Hoey's criterial question to identify evaluation (1983: 47). Sentence 2 merely seems to elaborate sentence 1, by giving further information concerning the identified protagonist. What is clear is that, as yet, no problem has been overtly signalled.

Sentence 3

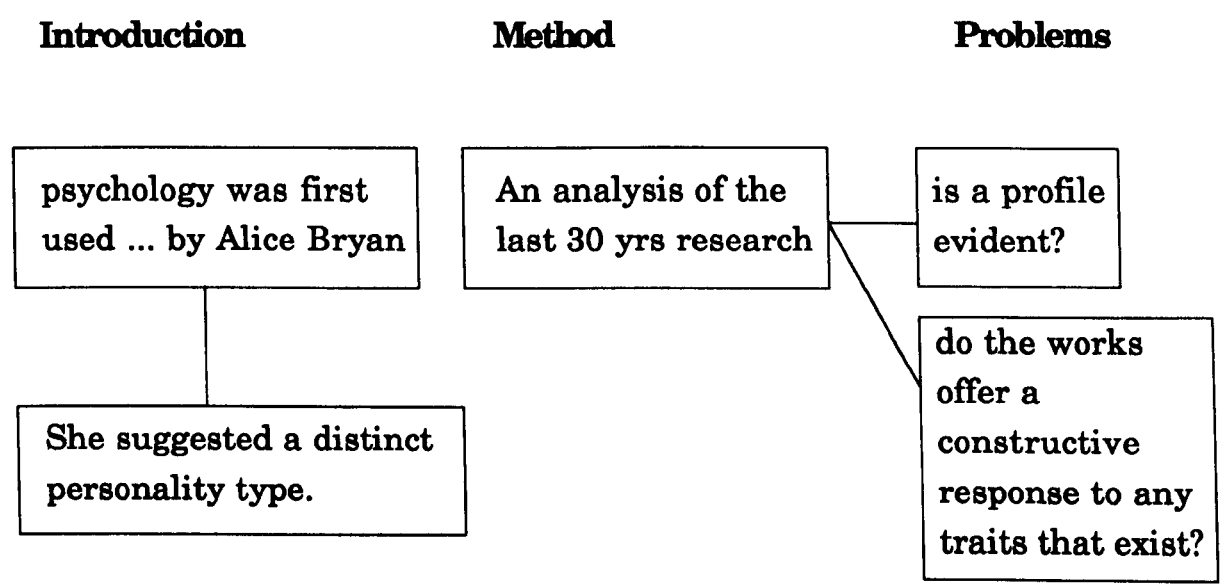
An analysis of the last 30 yrs research will show if a profile is evident, as well as if the works offer a constructive response to any traits that exist.

The most striking feature of the third sentence is its curious change of tense. The absence of a reporting, or projecting, verb together with the shift in tense make sentence 3 appear to be written using free indirect discourse (McHale 1978). According to the analyst's intuitions, this has a disorienting effect, perhaps because one of the functions of free indirect discourse is to introduce a 'plurality of speakers and attitudes', which in turn contributes to the 'semantic density of the text' (Rimmon-Kenan 1983: 114). In other words, it is

not clear whose voice is represented: the original author's or the abstractor's.

There does seem to be information in sentence 3 which is useful, but it is signalled in a less than transparent way. Looking at sentence 3 slightly more closely, there is material buried in the two conditional 'if' clauses which refer to two questions. Providing answers to these two questions could be construed as the *problem* the original article is meant to address. The first few words of the sentence, 'An analysis of the last 30 yrs research will show ...', suggest that the analysis constitutes the *method* by which the problem is to be attacked, but, again, the signalling is somewhat opaque.

The structure of the abstract so far can be diagrammed as follows:



At this stage it may be useful to compare and contrast this abstract with Distinct Personality Type, abstract D:

<u>Version C</u> (less successful)	<u>Version D</u> (more successful)
Problems first introduced in Sentence 3 (with inexplicit signalling)	Problems first introduced in sentence 1 (with explicit signalling)
Method first introduced in Sentence 3 (with inexplicit signalling)	Methods first introduced in sentence 1 (with explicit signalling)
Method precede Problems	Problems precede Methods

These differences may in part explain why version D is preferred over version C.

Sentences 4 to 8

Libraries do have an image which is dull and uninspiring, they are staffed by "fussy old women of either sex" and used by "unfortunate characters". This image is encouraged by media and other misrepresentation. Librarians image ambiguous as both severe and diffident. Psychologic studies are concerned with personality which is subjective, the one area of personality most used in research in this area is trait. It's an unfortunate choice as it assumes results from general behavioural patterns without taking into account time, personal experience and place.

How can one incorporate these sentences into the above diagram? Sentence 4 starts off what will be claimed to be the middle section of the abstract. According to the analyst's intuitions, the middle section (from sentence 4 to sentence 8) contains very general, background information, all of which is realised using the present tense. There is little explicit signalling present, which makes the generic structure unclear, and once again there is an ambiguity concerning authorial voice.

These sentences are interesting in that there is a good deal of cohesion linking successive sentences: note the carefully organised Thematic Progression ('TP'; see for example Danes 1974): simple linear TP linking sentences 4 and 5; TP with constant theme linking sentences 5 and 6; and simple linear TP linking the clauses in sentence 7 and linking sentence 7 to

sentence 8. In spite of the cohesion, however, it is difficult to see how these sentences contribute to the overall structure of the abstract.

Sentence 9

This area of personality, trait, has been used in research by a number of people including Douglass, McDermouth, and McMahon's who have achieved results which vary considerably.

This sentence changes the pace of the abstract once again, and moves from the general to the particular. Sentence 9 is almost a second introductory section, very similar to the first sentence (note the repeated verb 'used' and mention of further researchers in the field), except that it does contain additional evaluative material at the end of the sentence. 'Results' also feature in the following sentence:

Sentence 10

Very few surveys showed similar results.

Part of the difficulty, according to the analyst's intuitions at least, is that the abstractor provides insufficient signalling to disambiguate between the two different types of 'result' reported: the results of the author's review versus the chronologically prior results of the various studies the author consulted for his review.

The following three sentences contain a mixture of results and conclusions:

Sentences 11 - 13

There were problems with the survey both with standards - what is a "normal" score and with the numbers and methodology involved. The effects the librarians traits had on the organisation of a library were equally inconclusive. These surveys have brought the use of psychology research in this area into doubt.

As concluding material, these sentences are unsuccessful primarily because the basic research questions were not sufficiently well spelt out earlier in the abstract, so readers cannot be expected to know whether sentences 11 - 13 constitute satisfactory answers. Although version C 'loses out because of poor conclusion' (judge 1), there is a sense in which this was inevitable, because the aims and objectives of the review were not made clear at the beginning of the

abstract.

Once again, it is useful to contrast the conclusions from this abstract with those from Distinct Personality Type, abstract D. Version D featured a recognisable symmetry between its conclusions section and the problems section:

Conclusion 1 (sentence 6) answers Problem 1 (sentence 1);

Conclusion 2 (sentence 7) answers Problems 2 (sentence 2) and 4 (sentence 4);

Conclusion 3 (sentence 8) also answers Problem 1 (sentence 1).

This symmetry was entirely lacking in version C: if the two problems of C are in fact 'is a profile evident?' and 'do the works offer a constructive response to any traits that exist?', then there is nothing in sentences 11 - 13 which can be taken as an answer to these questions.

Perhaps this lack of direction in part prompted the following criticism of C: 'No clear plan/progression evident' (judge 8). Certainly, the structure in C is much less apparent.

9.7 Discussion

The two analyses above can be considered to be in some ways successful, and in other ways unsuccessful.

They have been successful in so far as it has been possible to demonstrate a certain correspondence between informants' intuitions and the relative ease of analysing the texts in terms of their generic structure.

However, the analysis has been less than wholly satisfactory for a number of reasons. The analyses offered above account for the analyst's intuitions only; no claims are made concerning their generality. In other words, the *reliability* of the technique cannot be guaranteed. This is because it is harder to provide an algorithmic characterisation of the process by which generic structure is recovered from texts than it is for virtually any other type of linguistic description. This is not to suggest that analysing texts reliably in terms of their generic structure will never be possible; however, recovering precise, reliable internal measures in order to test hypotheses relating linguistic features to judges' preferences would necessitate a reworking of the present framework on a scale more appropriate to an entire Ph.D., rather than a small pilot analysis.

The larger-scale investigations reported on in the remainder of Part Four, on the other hand, make use of linguistic tools which are better understood, are more reliable, and can provide more sensitive measures.

9.8 Conclusions

The following conclusions are to be drawn from this chapter:

- The analysis of generic structure has a certain intuitive appeal, in that it provides a relatively course-grained means of explaining why some abstracts are perceived as being more structured than others.

- Graetz's framework:

Problem - Methods - Results - Conclusions

and that of Hill et al:

Introduction (serving to point out the problems with previously held theory) -
Procedure (consisting of Methods and Results) - Conclusions

were found to be accurate in predicting which elements of generic structure have to be present in order for an abstract to be perceived to be successful.

- Further, there is evidence to show that the ordering of these elements is important: abstract C in which 'Method' preceded 'Problems' was perceived to be less successful than abstract D in which 'Problems' preceded 'Methods'.

- Also, it is not sufficient for an abstract to be structured; the signalling of the structure must be made clear. Abstract C suffers from a lack of clear signalling. Not surprisingly, it is particularly hard to apply the framework to texts which lack, fail to make clear, or inappropriately signal, their underlying structure.

- This type of analysis could not be taken further, principally because, currently, the framework cannot always be applied with the degree of reliability required for precise hypothesis testing.

10

Analysis of Lexical Texture

10.1 Introduction

This chapter investigates certain lexical aspects of the abstracts and the bearing these may have on informants' judgements. The term 'lexical texture' is taken to comprise the two concepts of lexical density and lexical variation, and should not be confused with the sense of the word 'texture' as used by Halliday and Hasan 1976, for example. Their conception of texture will be examined in later chapters: Chapter 11, which builds on the distinction between lexical and grammatical words discussed in this chapter, examines cohesive features, while Chapter 13, on choice of Themes, concentrates on more structural means of engendering texture.

The idea of lexical texture affecting judges' perceptions of text quality is due to Linnarud (1975, 1977), and is discussed in the next section, where separate consideration is given to lexical density (10.2.1) and to lexical variation (10.2.2). Two basic hypotheses are then stated (10.3), one relating to each concept.

Section 10.4 describes three decisions that were taken prior to the formal testing of the hypotheses, the first two of which (10.4.1 and 10.4.2) also have relevance for further chapters in Part Four. The third (10.4.3) details how quantitative measures were derived to test the two hypotheses.

Results are reported in section 10.5, where tables of scores are presented alongside each hypothesis. The results are discussed in section 10.6; finally, brief conclusions are given in 10.7.

10.2 Lexical Texture as a Determinant of Success

In the qualitative data, issues of readability and the differences between 'concise' and 'wordy' styles of writing are commented upon by the judges in

each of the four sets of abstracts, but they are most prevalent in the first set, which contains the five abstracts which summarise the article describing the plight of the library service in Tanzania:

concerning version A: 'quality of the writing not too good' (judge 2);

concerning version B: 'very concise, well expressed' (judge 7);

concerning version C: 'too wordy' (judges 1, 2, 8);
'well constructed and clear' (judge 8).

concerning version D: 'badly written, unclear' (judge 2);
'I didn't like the style of D very much - too chatty possibly' (judge 3);
'fragmented and uncohesive. Suffers from omissions ... and verbosity' (judge 8).

concerning version E: 'well written, clear and understandable' (judge 2);
'easy to read' (judge 3);
'easier to read [than D], with flowing concepts' (judge 6);
'well constructed on the whole ... verbosity detract[s]' (judge 8).

This chapter investigates the claim that an analysis of what is termed a text's 'lexical texture' (Linnarud 1975) can help explain such observations, more especially by predicting which of a number of different texts will be considered the most readable (Coleman 1971), and by correlating certain measures of lexical texture against the overall perceived quality of different written texts (Linnarud 1975).

Following Linnarud, two types of lexical measurement are made for every text, which together are assumed to provide an indication of its lexical texture. These are: *lexical density* (discussed in Halliday 1987, and more comprehensively in 1989); and *lexical variation* (see for example Linnarud 1975). These concepts will be considered in turn.

10.2.1 Lexical Density

Lexical density (LD) has to do with the relative numbers of lexical and grammatical words contained in a text. On first sight it would therefore appear to be a relatively straightforward notion. In practice, however, differentiating between the two kinds of word is much more problematic than most analysts would have their readers believe.

Stubbs 1986 provides a number of tests for the two types of word, but the most fundamental difference seems to be that lexical words form an open class, whereas grammatical words form a closed class (Halliday 1989: 61, Quirk et al 1972: 44 - 47). Although Lyons (1968: 435 - 436) claims that this is the most satisfactory criterion, there remain a large number of non-trivial problems to be solved if analysts are ever to have a wholly reliable means of differentiation; Parsons (forthcoming) gives some useful comments on the different possible ways forward.

Notwithstanding the problems, LD, which Ure (1971: 445) has described as the 'proportional occurrence of lexis' in a text (and is to be expressed as a percentage; see section 10.4.3 below), has been put to a variety of different uses.

In cognitive psychology, LD's effect has been studied both on sentence comprehension (Backman 1978) and on memory retention, often with seemingly contradictory findings: Perfetti 1969, for example, found that speakers are generally less able to recall material with a high LD than they are material with a low LD, while Murdock and Desmond 1975 concluded that they are more able to comprehend material with higher LD.

In text-based linguistics, on the other hand, Stubbs (1986: 33) says 'One important reason for studying L and G words is that their relative frequency in a text differs considerably according to various features of the context in which the text is produced'. Different features of the context investigated so far include: Ure's pioneering work on differences of mode (1971), in which she discovered a strong tendency for spoken texts to have LDs of less than 40%, written texts scoring above 40%; Zora and Johns-Lewis' investigation across two different text-types, interview and conversation, in which they concluded (1989: 97) that postgraduates' levels of LD rose when being interviewed, possibly 'to match some perceived characteristic of the interview situation'; and Linnarud's work (1975 and 1977), which compared and contrasted LD scores for Swedish and native English speakers both writing English, and attempted to relate these to judges' evaluations of the quality of the written work.

The exact means of calculating LD adopted in this study is detailed in section 10.4.3 below.

10.2.2 Lexical Variation

Lexical Variation (LV), sometimes called 'Lexical Diversity' (see for example Bradac et al 1977), has to do with the type/token ratio (Kucera and Francis 1970: 356) of the lexical words in a text. In other words, if writers choose to repeat numbers of lexical words many times, then their texts will score low on LV, whereas writers who are careful to vary their lexis (either because they are constantly referring to different entities, or because they employ synonyms or other sense relations) will produce texts that receive high LV scores.

It was considered important to examine LV in addition to LD, since Linnarud has written that LV 'may well be a more important factor in influencing the evaluator favourably than LD' (Linnarud 1977: 90). LV was thought to be particularly relevant for the abstracting context, in which issues of verbal economy are generally assumed to be very much to the fore.

The calculation of LV is given in section 10.4.3 below.

10.3 Statement of Hypotheses

If, as Bradac et al (1977: 274) claim, 'a high-density message contains more semantic information than does a low-density message', then it may be supposed that the better abstracts should have a relatively high LD, since abstract writers are typically expected to cram as much information into as small a space as possible, thereby maximising the information available to readers without making undue demand on their time.

But is it sufficient merely to have a large amount of information?; should abstract writers not include large amounts of *different* information in order to broaden their coverage of the source material? If this second premise is true, then it may be supposed that the better abstracts should have a relatively high LV. Another reason why one would expect the better abstracts to have higher LV scores is that *stylistic* differences will also engender greater lexical variation (see section 10.2.2 above). These suppositions are in keeping with Linnarud's initial assumptions; she hypothesised (1975: 50) that 'essays which

have both a high LD and a high lexical variation, combined with a lack of error, should make a better impression on the reader than one with a low value in one or both'.

The two hypotheses to be investigated in the current chapter can be formally stated as:

- **Hypothesis 1:**

Abstracts with higher LD will be generally preferred over abstracts with lower LD.

- **Hypothesis 2:**

Abstracts with higher LV will be generally preferred over abstracts with lower LV.

The means of testing these hypotheses are described in the next section, and the results of the testing are reported in section 10.5 below.

10.4 Testing the Hypotheses

This section describes three decisions that were taken prior to the hypothesis testing. The last of these (10.4.3) concerns the way scores for the two lexical measures were derived and calculated for each of the seventeen Information Science abstracts; these decisions are specific to this chapter. The first two decisions (10.4.1 and 10.4.2), however, concern not only this chapter, but also all the chapters in Part Four which attempt a reconciliation between the linguistic analysis on the one hand and judges' rankings on the other. The first of these (10.4.1) was motivated by the need to *reduce* the quantitative data elicited from the judges to facilitate hypothesis testing. The second (10.4.2) concerns the procedures used to compare the results of the linguistic analysis with the reduced rank scores.

10.4.1 Reducing the Judges' Rankings to more Manageable Size

Chapter 8 above provides a catalogue of the data collected throughout the course of the study, both qualitative and quantitative. Quantitative data

normally takes precedence over the purely qualitative at the hypothesis testing stage in a study such as this; consequently it is the judges' ranks that are of primary concern.

For the Tanzanian set, for example, these were as follows:

Abstract Set: Tanzania

Abstract Version	Judges							
	J1	J2	J3	J4	J5	J6	J7	J8
A	5	2	1	1	2	1	5	3
B	1	5	3	2	1	2	3	4
C	3	3	4	3	3	3	2	5
D	2	1	2	4	4	4	1	1
E	4	4	5	5	5	5	4	2

These figures are to be interpreted as follows: the first judge (J1), for example, considered version A to be the best abstract of the set of five, it therefore receives the top mark of 5. The second judge (J2), however, thought the same abstract was the second worst of the set, it therefore receives a score of 2.

Also in Chapter 8, the degree of agreement between judges' preferences was calculated using the Kendall Coefficient of Concordance, where a *w* value of 0 can be taken to mean that judges are allocating their rank scores randomly, whereas a value of 1 signifies complete inter-judge agreement. It was established that the scores for the four sets of Information Science abstracts were as follows:

Abstract Set	W
Distinct Personality Type	0.597
Tanzania	0.241
Vickery and Vickery, Chapter 1	0.531
Vickery and Vickery, Chapter 2	0.109

The fact that there is such a wide range of values, from a comparatively large amount of agreement (Distinct Personality Type), to almost complete disagreement (Vickery and Vickery, Chapter 2) is a very interesting finding, but presents certain practical difficulties.

Because judges differ so much between themselves, each hypothesis should ideally be investigated in relation to each judge. This would then mean that

each hypothesis would have eight separate sets of results, each one quite possibly being falsified for some judges, but supported for others. This could then be used to derive clusters of judges, where judges who had similar views concerning the importance of particular linguistic features would be grouped together.

However, as there is a relatively large number of hypotheses to be investigated in this thesis anyway, some form of data reduction was thought necessary. The decision was taken therefore to consider only the *overall perceived quality* of the abstracts throughout the analysis phase of the study. The overall perceived quality was arrived at as follows: the scores derived from the eight judge's ranks were summed to give a total score for each abstract.

Abstract Set: Tanzania

Abstract Version	Judges								Total Score	Position
	J1	J2	J3	J4	J5	J6	J7	J8		
A	5	2	1	1	2	1	5	3	20	4
B	1	5	3	2	1	2	3	4	21	3
C	3	3	4	3	3	3	2	5	26	2
D	2	1	2	4	4	4	1	1	19	5
E	4	4	5	5	5	5	4	2	34	1

These total scores were then ranked to give an overall position. For the Tanzanian set therefore, version E, with a total score of 34, was considered the best overall, version C second best, with only one point to separate versions B, A and D. It is also worth noting that E and C are the abstract versions over which there seems to be the most agreement, as shown by the similar values in each of the two rows of figures. The overall preferred ordering, E-C-B-A-D in this case, will be referred to in what follows as the *observed rank*; the next section discusses the means of comparing this ordering with that predicted by the various hypotheses.

10.4.2 Correlating Observed and Expected Ranks

The ordering predicted by a particular hypothesis will be referred to in what follows as the *expected rank*. The aim of the hypothesis testing, then, is to

assess the match between the expected and observed ranks, in order to show how well each of the hypotheses predicts the judges' attitudes, to see how well the theory fits the data, in other words.

Traditional, nonparametric measures of correlation (such as Spearman's, or Kendall's, rank correlation coefficients, or Pearson's product-moment correlation coefficient, see Siegel 1956), which would have been the normal means of assessing such a match, are unreliable however when the number of cases is small.

Although each judge was asked to assess a total of seventeen Information Science abstracts (enough data for such measures), the abstracts were presented to the informants in four sets, two sets of five abstracts, one of four, and one of three. These four sets pertain to the four *different* source texts, and so it would be inappropriate to assume that this design constitutes a homogenous set of seventeen cases, all ranked in respect of each other.

It is not legitimate therefore to use the type of statistical technique mentioned above, given the nature and extent of the numerical data available. This is not to imply that no attempt will be made to show how the expected ranks relate to the observed ranks; this will be done, but not by sophisticated statistical means (in 10.6, the concluding section below, a simple 'shifts away' measure is given to show how the observed and expected ranks can be crudely compared). It must be remembered that one of the basic aims of this study is to identify a number of promising-looking hypotheses which warrant further investigation, rather than to exhaustively test in rigorous statistical depth one or two hypotheses which have already been motivated by previous research. Firm statistical correlations between observed and expected ranks cannot therefore be expected to arise from this particular kind of exploratory study.

The next section describes how the expected ranks are derived.

10.4.3 Deriving Quantitative Lexical Measures

As mentioned above, differentiating reliably between lexical and grammatical words is both a difficult and time-consuming task. In this study it was decided to effect the analysis by means of computer. This has a number of advantages and a number of disadvantages. The advantages will be dealt with first.

First and foremost, computers are consistent. Calculating LD scores by hand is a laborious task, and errors are likely to creep in. Secondly, computers

save time, particularly when the calculations to be performed are numeric and routine. Thirdly, the fact that computers demand precise instructions forces analysts to provide algorithms, which in turn requires them to be fully explicit in their thinking. The computational method used to test the hypotheses in this chapter, for example, relies on an explicitly stated list of grammatical words (Appendix 10.1); any word not belonging in that list is thought to be lexical. The explicit statement of this list, together with the algorithm which derives the various scores (Appendix 10.2), means that the results described in this chapter are replicable by other researchers. This is not always true of LD studies reported in the literature, many of which do not provide an account of how the calculations were derived.

This is a compelling set of advantages, but it must be remembered that the use of computers to calculate LD does have one major disadvantage (see, for example, the warnings in Stubbs 1986). This disadvantage stems from the fact that some words are ambiguous; they can be either lexical or grammatical depending on how they are being used:

- 1) John is going to the pictures;
- 2) John is a teacher.

Most researchers would agree that the word 'is' in 1) above is grammatical, but lexical in 2). Currently the computer has no means of disambiguating such words; furnishing it with the additional contextual information that would be required would be a substantial undertaking, one which is beyond the scope of the present chapter.

The computer program written to test the hypotheses stated in this chapter was adapted from the algorithm presented in Stubbs 1986 (the revised algorithm is shown in Appendix 10.2). Very simply, each word in the input text is checked to see if it is in a previously-stored list of grammatical words. If it is, it is assumed to be grammatical; if not, it is assumed to be lexical.

Zora and Johns-Lewis (1989) say Stubbs' list of grammatical words is not exhaustive, and make a number of suggestions for inclusion. The list provided by Butler (1985b: 219 - 220) contains these and further additions, and so this was used in the present study (this is reproduced as Appendix 10.1).

Algorithms such as that used in this study merely perform tests of orthographic equality, and make no reference to syntactic or semantic information; the disadvantage of such simple computational procedures is that input words which can be both lexical and grammatical cannot reliably be

disambiguated, and so inaccuracies may result. The problem is most usually handled in terms of statistical likelihood, where each word is assumed to belong to one word class only: so, for example, 'can' is thought to be grammatical, since its lexical use (to denote a receptacle) is altogether more unusual.

The actual formulae used for calculating the LD and LV percentages (Linnarud 1975: 46) are as follows:

$$\begin{aligned} \text{LD} &= \frac{\text{Number of Lexical Words}}{\text{Total Number of Words}} \times \frac{100}{1} \\ \text{LV} &= \frac{\text{Number of Different Lexical Words}}{\text{Total Number of Lexical Words}} \times \frac{100}{1} \end{aligned}$$

The first of these makes more explicit Ure's original formulation of LD. This accords with Halliday's definition of LD (1987: 60) as 'the proportion of lexical items (content words) to the total discourse'. However, he continues 'It can be measured in various ways: the ratio of lexical items either to total running words or to some higher grammatical unit, most obviously the clause'. The formula for LD stated above captures Halliday's 'running words' formulation, but in the more extended account (1989: 63 - 67), he makes clear his preference for using the measure which makes appeal to the higher grammatical unit: 'words are not packed inside other words; they are packaged in larger grammatical units - sentences, and their component parts. It is this packaging into larger grammatical structures that *really* determines the informational density of a passage of text' (1989: 66, emphasis added). Slightly later, however, Halliday states that it should be measured as the number of lexical items per clause, and that 'No account need be taken, for purposes of this particular measurement, of the number and organisation of clauses in the sentence (clause complex)' (1989: 67).

Without more justification this is a somewhat curious statement. There appears to be no a priori reason why the clausal measure of lexical density should necessarily be more revealing than that according to clause complex. In fact, the following discussion will argue that it is improper to incorporate *any* measure of grammatical complexity into what is supposed to be a measure of lexical complexity, the most fundamental objection being that a grammatically normalised LD score can introduce bias which has no bearing on the true lexicality of the text. The danger is that researchers can be misled

into believing a text has high LD, when in reality it is lexically sparse, the high score being a function of its grammatical complexity.

Halliday explains the relationship that obtains between grammatical intricacy (see Chapter 12 below) and lexical density in terms of the cline between speech and writing, noting that, most typically, writing tends to be lexically dense and grammatically simple, whereas speech tends to be lexically sparse and grammatically intricate. The fact that this appears to be such a strong tendency, however, should not be construed as meaning that the two variables necessarily co-vary. In fact, the position adopted in this study is that grammatical intricacy and lexical density should be regarded as *independent* variables, and, as such, should be investigated separately in the attempt to factor out each of their individual contributions to the perceived quality of the discourse. Consequently, only Ure's original ratio measure (corresponding to Halliday's running words formulation) is used in this chapter. Grammatical intricacy will be investigated in Chapter 12.

LV, the second kind of measure investigated in this chapter, characterises the lexical type/token ratio of a text. Stubbs' algorithm was not designed to calculate LV, so it had to be adapted to incorporate a measure of *different* lexical items. Appendix 10.2 shows how a table of 'lexwords' was appended and checked as the program proceeded, such that it contained all and only the different lexical items present in the input text.

As with LD, careful consideration should be given to the LV formula stated above. Linnarud 1975 points out that the length of the text plays an important part: 'As the essays are of different lengths, the figures giving the variation for each are not strictly comparable. The presumption must be that a high figure is easier to gain in a short essay than in a long' (1975: 50). This problem will be returned to in section 10.5.2 below.

Finally, the fact that the computer's calculation of LD and LV is reliable is no guarantee of its validity. In other words, the computer performs all and only the operations it is programmed to, but care must be taken to ensure that its instructions are genuinely what is required. The three shortest abstracts were checked by hand, therefore, to ensure that the computer was behaving as predicted. The manual check of the LD of the shortest abstract, Tanzania version B, is shown in Appendix 10.3.

10.5 Results of Testing the Hypotheses

In this section the two hypotheses are investigated separately (sections 10.5.1 and 10.5.2), using ranks representing the overall perceived quality of the abstracts, as discussed above. In each case, the hypothesis is stated once again and followed by tables of results.

10.5.1 Hypothesis 1 Results

The first hypothesis stated that abstracts with higher LD would be generally preferred over abstracts with lower LD. The LD scores for the four sets of abstracts are presented on the following two pages:

Lexical Density Results

Abstract Set: Tanzania

Abstract Version:	A	B	C	D	E
Total number of words (T):	198	103	237	274	277
Total number of grammatical words (G):	69	40	101	102	113
Total number of lexical words (L):	129	63	136	172	164
Lexical Density:	65.1%	61.1%	57.4%	62.8%	59.2%

Lexical Density Results

Abstract Set: Distinct Personality Type

Abstract Version:	A	B	C	D	E
Total number of words (T):	651	323	230	191	127
Total number of grammatical words (G):	222	132	92	83	58
Total number of lexical words (L):	429	191	138	108	69
Lexical Density:	65.9%	59.1%	60%	56.5%	54.3%

Lexical Density Results

Abstract Set: Vickery and Vickery, Chapter 1

Abstract Version:	A	B	C	D
Total number of words (T):	268	272	238	471
Total number of grammatical words (G):	103	109	82	184
Total number of lexical words (L):	165	163	156	287
Lexical Density:	61.6%	59.9%	65.5%	60.9%

Lexical Density Results

Abstract Set: Vickery and Vickery, Chapter 2

Abstract Version:	A	B	C
Total number of words (T):	199	265	111
Total number of grammatical words (G):	93	109	48
Total number of lexical words (L):	106	156	63
Lexical Density:	53.3%	58.9%	56.8%

For the Tanzanian set, the observed and expected ranks are as follows:

Abstract Version	A	B	C	D	E
Preference Score	20	21	26	19	34
% LD	65.1	61.1	57.4	62.8	59.2
Observed Rank	E	C	B	A	D
Expected Rank	A	D	B	E	C

In this table the observed rank is obtained by ranking the overall preference scores, while the expected rank is obtained by ranking the LD scores according to the order predicted by the hypothesis. In this case the hypothesis is most certainly falsified: the two abstracts considered the best by the judges, versions E and C, have the lowest LD values, while those considered the two worst, versions A and D, have the highest LD values. Turning to the qualitative comments very briefly, three separate judges used the lexicalisation 'wordy' of version C, which is interesting and unexpected, since C has the lowest LD score. 'Wordiness' will be returned to below.

For the Distinct Personality Type set, the observed and expected ranks are as follows:

Abstract Version	A	B	C	D	E
Preference Score	8	31	22	32	27
% LD	65.9	59.1	60	56.5	54.3
Observed Rank	D	B	E	C	A
Expected Rank	A	C	B	D	E

Again, the hypothesis is not supported. Further, there seems to be evidence that the reverse is true. Judges seem to disprefer those abstracts with high LD; once again, the two densest abstracts are the two least preferred by judges.

For the Vickery and Vickery, Chapter 1 set, the observed and expected ranks are as follows:

Abstract Version	A	B	C	D
Preference Score	9	23	22	26
% LD	61.6	59.9	65.5	60.9
Observed Rank	D	B	C	A
Expected Rank	C	A	D	B

The fact that abstract version A received such a small overall preference score is not well explained by the LD scores. Thus it must be assumed that the cause of the judges' dislike of version A pertains to a different linguistic feature altogether. Once again, however, the general tendency for the hypothesis to be reversed is evident. Interestingly, the first judge remarked that version C was 'a bit wordy', an observation which accords more with its high LD score.

'Wordiness' was also a criticism of version B of the Vickery and Vickery, Chapter 2 set:

Abstract Version	A	B	C
Preference Score	17	13	18
% LD	53.3	58.9	56.8
Observed Rank	C	A	B
Expected Rank	B	C	A

Again the hypothesis is falsified. B has the highest lexical density, perhaps explaining its perceived 'wordiness'. Pretheoretically, perceptions of wordiness would be expected to correlate with high LD scores, as witnessed in the last two sets presented above, but the fact that three judges complained that Tanzania version C was 'wordy' is strange considering it is the least dense of the five. It would seem then that some facets of wordiness are associated with high LD scores, but there are others which seem to be independent. Grammatical intricacy, which is often claimed to be the counterpart of lexical density, may explain those other facets of wordiness and will be investigated in Chapter 12 below.

10.5.2 Hypothesis 2 Results

The second hypothesis stated that abstracts with higher LV would be generally preferred over abstracts with lower LV. The LV scores for the four sets of abstracts are presented on the following two pages:

Lexical Variation Results

Abstract Set: Tanzania

Abstract Version:	A	B	C	D	E
Total number of words (T):	198	103	237	274	277
Total number of lexical words (L):	129	63	136	172	164
Number of different lexical words (D):	105	56	116	126	118
Lexical Variation:	81.4%	88.9%	85.3%	73.3%	72%

Lexical Variation Results

Abstract Set: Distinct Personality Type

Abstract Version:	A	B	C	D	E
Total number of words (T):	651	323	230	191	127
Total number of lexical words (L):	429	191	138	108	69
Number of different lexical words (D):	287	141	102	76	55
Lexical Variation:	66.9%	73.8%	73.9%	70.4%	79.7%

Lexical Variation Results

Abstract Set: Vickery and Vickery, Chapter 1

Abstract Version:	A	B	C	D
Total number of words (T):	268	272	238	471
Total number of lexical words (L):	165	163	156	287
Number of different lexical words (D):	110	107	111	124
Lexical Variation:	66.7%	65.6%	71.1%	43.2%

Lexical Variation Results

Abstract Set: Vickery and Vickery, Chapter 2

Abstract Version:	A	B	C
Total number of words (T):	199	265	111
Total number of lexical words (L):	106	156	63
Number of different lexical words (D):	72	100	45
Lexical Variation:	67.9%	64.1%	71.4%

As mentioned above, care must be taken in the interpretation of the LV figures, since the overall length of the abstracts can skew the results. Indeed, as the following table shows, LV scores seem to be more determined by length than anything else (circles denote the shortest abstract of the set, squares the longest):

Abstract Set	LV Scores				
	A	B	C	D	E
Tanzania	81.4%	88.9%	85.3%	73.3%	72%
Distinct Personality	66.9%	73.8%	73.9%	70.4%	79.7%
V&V Chapter 1	66.7%	65.6%	71.1%	43.2%	
V&V Chapter 2	67.9%	64.1%	71.4%		

Thus the calculation for LV given in section 10.4.3 above *always* results in the shortest texts receiving the highest lexical variation rating, with the longest texts receiving the lowest lexical variation rating. This is clearly unsatisfactory, and so the hypotheses were retested using the following formula, which takes the overall length of the texts into account:

LV

=

Number of Different Lexical Words

Total Number of Words

x

100

1

This is different from the formula given in section 10.4.3 above, in which the denominator was the Total Number of Lexical Words; here the denominator is simply the Total Number of Words in the input text. The LV percentages calculated according to this new formula are as follows:

Revised Lexical Variation Results

Abstract Set	LV Scores				
	A	B	C	D	E
Tanzania	53%	54.4%	48.9%	46%	42.6%
Distinct Personality	44.1%	43.6%	44.3%	39.8%	43.3%
V&V Chapter 1	41%	39.3%	46.6%	26.3%	
V&V Chapter 2	36.2%	37.7%	40.5%		

The new set of observed and expected ranks are:

Tanzania: Observed Rank	E	C	B	A	D
Tanzania: Expected Rank	B	A	C	D	E
DPT: Observed Rank	D	B	E	C	A
DPT: Expected Rank	C	A	B	E	D
V&V1: Observed Rank	D	B	C	A	
V&V1: Expected Rank	C	A	B	D	
V&V2: Observed Rank	C	A	B		
V&V2: Expected Rank	C	B	A		

This last table shows that, in the first three sets, the abstract considered the best by the judges has in fact the lowest LV score of its competitors. It would seem that writing an abstract with many different lexical words guarantees greater perceived quality in only one of the four sets (V&V2); in two of the remaining three, the abstract with the highest LV is considered to be the second poorest. The hypothesis tested using the revised LV measure must therefore be considered to have been falsified.

It must be remembered that while this second LV measure is arguably more valid than the first, it still does not *solve* the problem of comparing texts of unequal length: Swales warns (1985: 5) 'A final but important point about type-token ratios is that only texts of the same number of words can be compared'.

Biber's formulation of this problem (1988: 239, emphasis added) is to say that 'the relation between the number of "types" (different lexical items) and the total number of words in a text is *not linear*'; in his own work he combats this problem by normalising for length, considering only the first 400 words of every text. Each set of abstracts could have been normalised by considering only the first n words of each abstract version, where n was the number of words in the shortest version of the set. However, this approach was not adopted in this study, because it would not have been proper to have compared the judges' rankings of version A of the Distinct Personality Type set, for example (ranks derived from their impressions of a 651 word-long abstract) with ranks expected by LV scores derived from the first 127 words of that abstract, 127 being the number of words in the shortest version of the DPT set, version E.

10.6 Discussion of the Results

The assumption underlying the hypotheses was that readers would appreciate having the maximum amount of new information presented to them in as economical a way possible. For the abstracting context this seems a very reasonable thing to expect.

However, the results do not bear out this assumption. The two hypotheses were falsified, but in such a way as to suggest that, for many cases, precisely the reverse situation seems to hold.

Hypothesis 1 stated that abstracts with higher LD would be generally preferred over abstracts with lower LD. This hypothesis was falsified. The

observed and expected ranks were as follows:

Tanzania: Observed Rank	E	C	B	A	D
Tanzania: Expected Rank	A	D	B	E	C
DPT: Observed Rank	D	B	E	C	A
DPT: Expected Rank	A	C	B	D	E
V&V1: Observed Rank	D	B	C	A	
V&V1: Expected Rank	C	A	D	B	
V&V2: Observed Rank	C	A	B		
V&V2: Expected Rank	B	C	A		

The LD scores for all the abstracts are well above Ure's 40% threshold, which is what would be expected of written texts; however, they seem to go against the common sense intuition that the better abstracts would be semantically 'heavier' and would therefore exhibit higher values of LD.

Hypothesis 2 stated that abstracts with higher LV would be generally preferred over abstracts with lower LV. This hypothesis was also falsified. The observed and expected ranks (using the second of the two LV formulae) were as follows:

Tanzania: Observed Rank	E	C	B	A	D
Tanzania: Expected Rank	B	A	C	D	E
DPT: Observed Rank	D	B	E	C	A
DPT: Expected Rank	C	A	B	E	D
V&V1: Observed Rank	D	B	C	A	
V&V1: Expected Rank	C	A	B	D	
V&V2: Observed Rank	C	A	B		
V&V2: Expected Rank	C	B	A		

The difference between observed and expected ranks can be quantified very roughly by a 'shifts away' score. When there is a one to one correlation between ranks, the columns line up perfectly, and nothing has to be shifted, as it were:

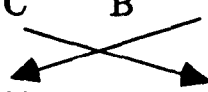
A D C B E

A D C B E

In the following scenario, however:

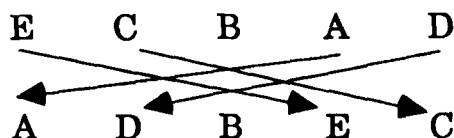
A D C B E

A D E B C



version C in the observed rank has been shifted forward two columns to match with version C in the expected rank, while version E has been shifted back two columns. It can therefore be said that the expected rank is four shifts away from the observed rank.

Similarly, the following ranks:



have a shifts away score of 12. 12 is the highest possible shifts away score for two sets of five items; 8 is the highest for two sets of four items; and 4 is the highest for two sets of three items.

Interestingly, the LD hypothesis was more impressively falsified than the LV hypothesis, as can be seen from the following shifts away score totals:

	LD	LV
Tanzania	12	10
Distinct Personality Type	12	12
Vickery and Vickery, Chapter 1	8	8
Vickery and Vickery, Chapter 2	4	2
	<hr/>	<hr/>
Shifts away totals	36	32

These figures indicate that the LD hypothesis could not have been more wrong, as it were, 36 being the maximum possible score.

Turning briefly to the interpretation of the findings from the two hypotheses, it is as if judges are less impressed by informationally-heavy abstracts, but respond more favourably to abstracts which are reader-friendly (a similar conclusion is reached in Chapter 12 below).

A cursory glance at the judges' verbal comments lends general support to this observation, but when remarks which contain the lexicalisation 'clear' or 'unclear' are collated, an interesting correspondence between the qualitative data and the LD scores is revealed.

Of the Tanzanian abstracts, version C, which judge 8 considered 'well constructed and clear', has the lowest LD, while version E, which judge 2 considered 'well written, clear and understandable' has the second lowest LD. D, on the other hand, the version with the second highest LD, was pronounced 'badly written, unclear' by the same judge.

Of the Distinct Personality Type abstracts, version C, which has the second highest LD, was considered 'not clear as to the salient points of the article' by judge 2, while version D, which 'seemed to extract key points and present them in a clear way', features the second lowest LD score.

Version B, the least lexically dense of the abstracts written for the first chapter of the Information Science book, according to judge 2, 'gives a clear picture with a logical progression in the argument', and according to judge 8, is 'not easily read, but its individual sentences are clear and well constructed'.

The results of the hypothesis testing would therefore seem to suggest that the judges generally prefer those abstracts which are more reader-friendly, and are prepared to tolerate some sacrifice in information content. Low levels of LD and LV seem to be more a feature of considerate abstract writing; high levels of LD and LV seem to be more a feature of those abstracts which are densely packed with information, but have associated processing costs in terms of reader comfort.

The next chapter continues the lexical theme by investigating the cohesion displayed by the various abstracts.

10.7 Conclusions

The following conclusions are to be drawn from this chapter:

- Lexical Density (LD) and Lexical Variation (LV) are relatively easy to quantify; the process can be carried out most reliably by means of a computer.
- However, care must be taken over the formulae used. Measures of LV in particular should be treated with caution when texts of different length are to be compared.
- The LD scores for all the abstracts are well above Ure's 40% threshold (1971), which is what would be expected of written texts.
- Because one would expect good quality abstracts to contain large amounts of different information expressed as economically as possible, it was hypothesised that the most preferred abstracts would be characterised by higher levels of LD and by higher levels of LV.
- In fact, both these hypotheses were falsified. In the case of LD especially, precisely the reverse seemed to hold: counter-intuitively, the better abstracts were characterised by lower levels of LD.

- A correspondence was noted between the LD scores and a number of the judges' qualitative comments. Abstracts with lower LD scores tend to be described as being 'clear'.

- It is suggested that low levels of LD and LV are more the mark of 'reader-friendly' abstract writing, whereas higher levels of LD and LV characterise abstracts which contain more information, but are correspondingly harder to process.

11

Analysis of Cohesion

11.1 Introduction

The Introduction to Part Four above stated that Halliday 1985 would be used as a convenient starting point for deciding which types of analysis could be investigated. Chapter 9 of Halliday 1985 might have been thought to be an appropriate departure point for an analysis of cohesive relations, but this chapter is to a very large extent a reworking of earlier ideas, most comprehensively detailed in Halliday and Hasan 1976. This work has been strongly criticised throughout the intervening period from a variety of different standpoints. Attacks have been made upon the theoretical adequacy of the framework: a common complaint is that the theory is altogether too heavily biased towards the surface text, and pays insufficient regard to the surrounding world, and to the cognitive aspects of the discourse (e.g. Huddleston 1978, Morgan and Sellner 1980, Carrell 1982 and Brown and Yule 1983); Noël 1989 notes that the categories postulated are poorly motivated, and that proposed text relations are seldom adequately defined.

Of perhaps more immediate relevance to the applied textlinguist, doubts have been expressed concerning both the usability and the usefulness of the system: Witte and Faigley 1981 were only able to demonstrate a slight correlation between the distribution of cohesive relations in college freshmen essays and their perceived quality; and, amongst others, Pritchard 1980 and Mosenthal and Tierney 1984 were forced to conclude that the number of cohesive ties in texts seems to have no direct bearing on how easy or difficult they are to understand.

However, although Neuner found that the distribution of the various types of lexical and grammatical ties did not help differentiate between good and bad texts, analyses based upon *cohesive chains* proved more useful, finding that 'the chains in good essays are denser, richer, and more interwoven than those in poorer essays' (1985: 10). Similarly Parsons 1990 reports encouraging

results using Hasan's more recent work on *chain interaction* (1984, 1985/1989b¹) to account for success judgements of texts written by native and overseas students.

Given these various points, Halliday and Hasan's original admonition not to expect cohesion to provide a measure of goodness for texts (1976: 328), and the fact that Hasan herself seems to have altered her position on the interdependence of cohesion and coherence (1985/1989b: 91, 1984: 190), cohesive harmony was felt to be the preferred means of carrying out a pilot analysis of cohesion.

In addition, Hartnett's work is briefly considered in section 10.3 below, since it is of applied pedagogical interest. Her doctoral dissertation (1980) posed the question, can cohesion be taught?, but resulted in only very limited improvement in the quality of students' writing. Adopting a more developmental perspective (1982), she demonstrated that progress could be made by charting certain *types* of ties, arguing that poorly rated essays typically lacked so-called 'manipulative' ties. In an article published in 1986, these ideas are recast to form a fully-fledged distinction between static and dynamic ties, the former serving to 'hold attention on a topic without necessarily manipulating it or changing it in any way', the latter to 'indicate how [an idea] develops, changes, or relates to something else' (1986: 145). Following suggestions made in this last paper, the distribution of these two different types of tie was piloted, to discover whether qualitatively different abstracts would exhibit different ratios of static : dynamic ties.

Because so many analytical problems were encountered in the pilot analysis of cohesion, it is proper, as was the case in Chapter 9 above, to preface the following discussion with a short statement pointing out what readers can and cannot expect to gain from reading further.

A number of problems are highlighted in what follows which should be of use to further researchers who may require to refine Hasan's framework. Also, a number of suggestions are made which may help to make the process by which identity chains are identified more explicit and thereby more reliable (see section 11.2.1.2 below). However, it was not possible to use the existing cohesive harmony framework to show why some abstracts were considered to be more successful than others, mainly because it was felt that Hasan's work does not yet constitute a sufficiently reliable tool for the purposes of precise hypothesis testing.

¹This article, originally written in 1980, appeared in a book published by the Deakin Press in 1985, and was subsequently reissued by the Oxford University Press in 1989. Page numbers in the text refer to this latest edition.

11.2 Cohesive Harmony

Hasan's later work (1984, 1985/1989b) was considered relevant to the present study, since it investigates linguistic phenomena which aid the production of coherent text. It supersedes Halliday and Hasan 1976 in that it more squarely addresses the relationship between cohesion and coherence through the attempt to reconcile a close textual analysis of a small number of young children's narratives with readers' judgements of coherence. In her two articles, Hasan presents a number of hypotheses which can be tested by transferring the framework she proposes to new and different texts, categorising lexical items according to the degree to which they interact with others, computing a number of different ratios, and then seeing how well these measures predict perceived levels of coherence. The basic premise is that it is not sufficient merely to have a number of similar lexical items in a text in order for it to be considered coherent, but rather that similar things must be said about other similar things. In this way the emphasis is shifted from chains *per se* to the extent to which, and to the means by which, they inter-relate.

In what follows an analysis of the cohesive harmony of Distinct Personality Type abstract version D is attempted, and problems described. The abstract is reproduced in full below and in Appendix 6.4.1. This abstract was chosen for two reasons: because it illustrates a number of interesting problems concerning the analysis of chain interaction; and because it has already been used in the pilot study of Generic Structure (see Chapter 9 above). Particular attention is paid to those areas in which it was deemed appropriate to depart slightly from the procedures suggested in the source literature.

11.2.1 Stages in the Analysis

There are four basic stages involved in the analysis: lexical rendering (Hasan 1985/1989b: 86 - 88); the formation of chains (1985/1989b: 89 - 90); the plotting of interactions (1985/1989b: 91 - 93); and the computation of various relevance measures or ratios (1985/1989b: 93 - 94), which, when considered in tandem with the overall cohesive linkage of the text, define the cohesive harmony of the discourse, which, in turn, is thought to provide an indication

of its general level of coherence. The first three of these four stages are detailed with respect to Distinct Personality Type, abstract D in the three subsections which follow (11.2.1.1 - 11.2.1.3).

11.2.1.1 Lexical Rendering

Hasan says that lexical rendering (1985/1989b: 87) chiefly involves two operations: the interpretation of grammatical cohesive relations (which entails the resolution of pronominal references and the recovery of substituted or ellipted elements), and the removal of any remaining non-content, or function, words (Ure 1971, Stubbs 1986).

Thus, according to Hasan, the lexical rendering of 'once upon a time there was a little girl' (clause complex 1 of text 5.1, 1985/1989b: 72) is 'little girl was' (1985/1989b: 87). This is a somewhat counter-intuitive rendering, because 'once' and 'time' have also been removed even though they are not function words (see Chapter 10 above). In fact, there appear to be a small number of inconsistencies in the way Hasan lexically renders her texts, so a simple computer program was written to remove function words more reliably using the list derived for the analysis of lexical texture above (see Appendix 10.1). The program also counted the frequency of the remaining lexical words, and identified their position by clause complex number.

This alternative lexical rendering differs from Hasan in the following respect: the alternative rendering process will exclude fewer lexical items than would Hasan's, because adverbials, for example, would survive the rendering. This was thought preferable since there appears to be no a priori reason why such elements should be excluded, and, in any event, the procedure advocated here has been automated to produce a more reliable analysis.

Distinct Personality Type abstract D is reproduced here in full:

1. The article sets out to answer the question, is the librarian a distinct personality type?, by a process of analysing the findings of psychological research conducted over the last 30 years, and a critical review of the methodologies used.
2. The author also addresses the question whether or not psychology has anything constructive to say about the library as an organization.

3. The popular stereotype is examined and the author concludes that, in popular imagination, the librarian is seen as both diffident and severe.
4. Using this hypothesis David Fisher reviews the major studies of the last three decades to ascertain how much fact there is in the stereotype, and how far the researchers have been affected by stereotypes.
5. In discussing personality the author points out that the word is rarely defined, and that most of the researchers assume an "unspoken" consensus of opinion of meaning.
6. David Fisher is highly critical of the research methods used, and concludes that on the evidence reviewed "it is not possible to state that the librarian is a distinct personality type".
7. "The utility of the whole psychological approach is put in doubt".
8. There is not a distinct library personality.

A lexically rendered version of Distinct Personality Type abstract D is therefore:

1. article sets answer question is librarian distinct personality type process analysing findings psychological research conducted last 30 years critical review methodologies used.
2. author also addresses question psychology has constructive say library organization.
3. popular stereotype examined author concludes popular imagination librarian seen diffident severe.
4. Using hypothesis David Fisher reviews major studies last three decades ascertain fact is stereotype far researchers have been affected stereotypes.
5. discussing personality author points word is rarely defined researchers assume unspoken consensus opinion meaning.
6. David Fisher is highly critical research methods used David Fisher concludes evidence reviewed possible state librarian distinct personality type.
7. utility whole psychological approach is put doubt.
8. is distinct library personality.

The underlining shows 'the interpretation of a grammatical cohesive device' (1985/1989b: 87): "David Fisher" is the elided grammatical subject of the second clause.

11.2.1.2 The Formation of Chains

Hasan's next step is the grouping together of lexical items into two kinds of chains. *Identity chains* are comprised of items which are co-referentially related (1984: 205) whereby each item 'refers to the same thing, event, or

whatever' (1985/1989b: 84). *Similarity chains*, on the other hand, are comprised of items which 'refer to non-identical members of the same class of things, events, etc., or to members of non-identical but related classes of things, events, etc.' (1985/1989b: 84).

It is to be expected that recognising similarity chains (SC's) will be somewhat less reliable than recognising identity chains (IC's), since sense relations, and, in the earlier paper, collocation are involved. These types of semantic link are problematic in that it is notoriously difficult to achieve agreement between different analysts (see below). And, while the recognition of identity chains is easier, there still remain a number of non-trivial problems.

Empirical investigation of the abstracts has revealed that the lexical items which form IC's are most usually nouns or adjectives; verbs *are* repeated, often in identical morphological form, but typically the actions referred to are *different* events (this also holds true for the metadiscoursal usage, "the author argues ... he argues", which are better treated as being verbally non-coreferential, since the projected material is different in each case). Consequently the approach taken here allows only nominals, defined by Halliday as all types of nouns, together with adjectives, and numerals (1985: 191), determiners having already been filtered out with the function words, to be candidates for inclusion in IC's. This additional constraint makes the analysis simpler, but it should be remembered that this modification (and all others made throughout the piloting) is designed to account *for this genre only*; no claims are made as to its suitability for other types of text.

Hasan does not provide a fully explicit description of the process by which IC's are constructed, but readers are invited to consider the following four stage algorithm which has been used in the present study:

- 1) Using the wordlist generated by the computer (see above), see which lexical items are repeated;
- 2) Check to see whether any sets of repeated lexical items can be collapsed on the grounds of co-referentiality. This is done to reduce the number of candidate IC chains. For example, in Distinct Personality Type version D, the fact that David Fisher *is* the author means that the two sets of referring expressions (the repeated lexical item "author" and the repeated mentions of "David Fisher") can be subsumed under the one IC;

- 3) Each set of repeated lexical items can then be considered to be a candidate IC;
- 4) Finally, check to see the lexical items in the candidate IC's really are co-referential.

Applying steps 1 to 3 of the above algorithm results in the following candidate IC's (these are shown in text order):

<i>lexical item</i>	<i>frequency</i>	<i>clause complex number</i>
question	2	1, 2
librarian	3	1, 3, 6
distinct	3	1, 6, 8
personality	4	1, 5, 6, 8
type	2	1, 6
psychological	2	1, 7
research	2	1, 6
[last 30 years/ last three decades	2	1, 4
critical	2	1, 6
[author	3	2, 3, 5
David Fisher	3	4, 6
library	2	2, 8
popular	2	3
researchers	2	4, 5
stereotype	2	3, 4

Notice that two sets of lexical items have been collapsed on the grounds of co-referentiality (shown by the brackets). These 14 candidate IC's can be diagrammed as shown below in Figure 11.1.

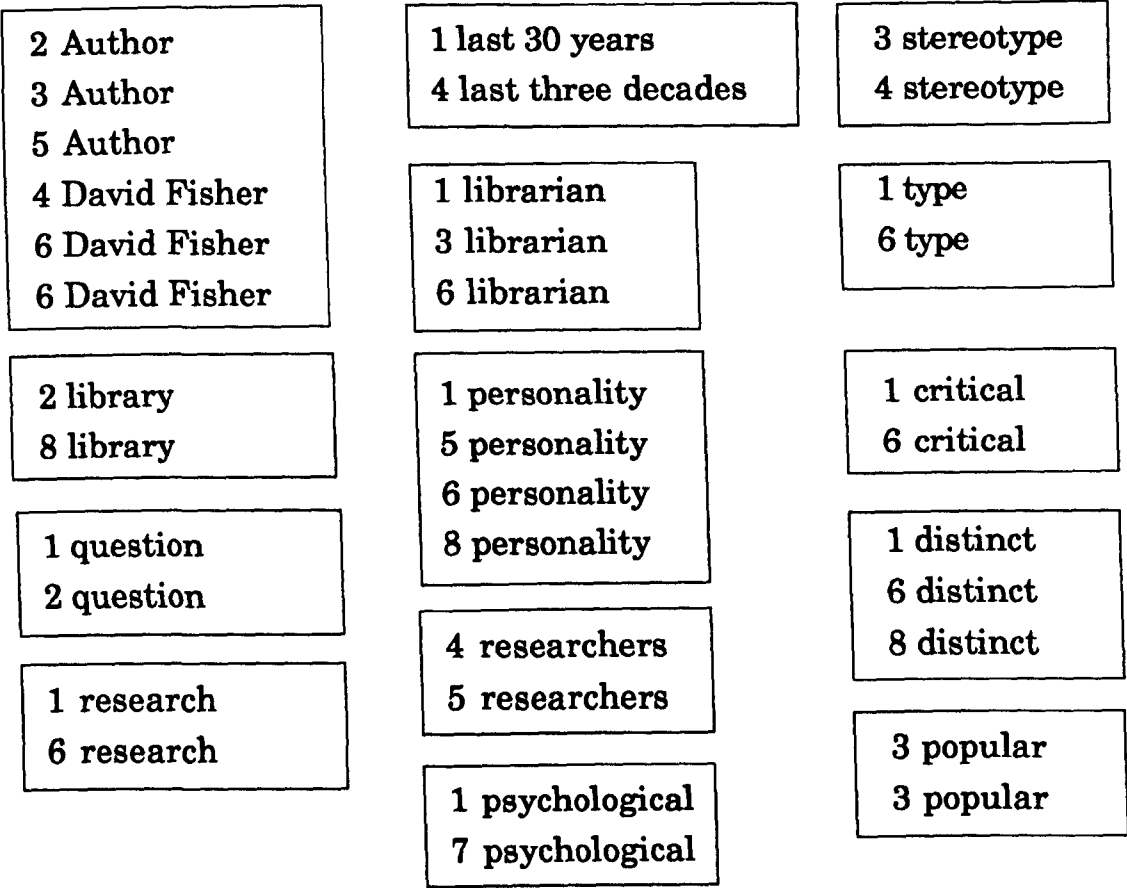


Figure 11.1

This diagram needs some modification, however, to reflect the truer picture. In other words, step 4 of the algorithm ("Finally, check to see the lexical items in the candidate IC's really are co-referential") must now be applied.

Although Hasan does not make the following point, the present study claims that *a single chain can only contain members of the same syntactic class*. This, it is submitted, is a useful, formal means of helping to ensure that only genuinely coreferential items are allowed to form IC's.

In the above diagram, however, some of the chains comprise members of different syntactic classes; a revision is therefore necessary if true coreferentiality is to be preserved. The "library" of clause complex 2 is a noun, whereas that of 8 is an noun modifier. In addition, the "research" of clause complex 1 is a noun, whereas that of 6 is a noun modifier. Similarly, the "personality" chain must be rewritten as two separate IC's: one for the noun modifier pair (clause complexes 1 and 6); and one for the noun pair (clause complexes 5 and 8). Also, the "question" pair constitutes a spurious IC, as inspection of the text reveals that the lexical item refers to two distinct

questions. Further, *adjectives that modify different headwords are assumed to be non-coreferential*; this stipulation allows the "distinct", "personality", and the "critical" clusters, but disallows "popular" and "psychological" ("popular" modifies "stereotype" but then "imagination" in clause complex 3; "psychological" modifies "research" in clause complex 1 but then it modifies "approach" in clause complex 7). A more accurate picture is therefore given by Figure 11.2.

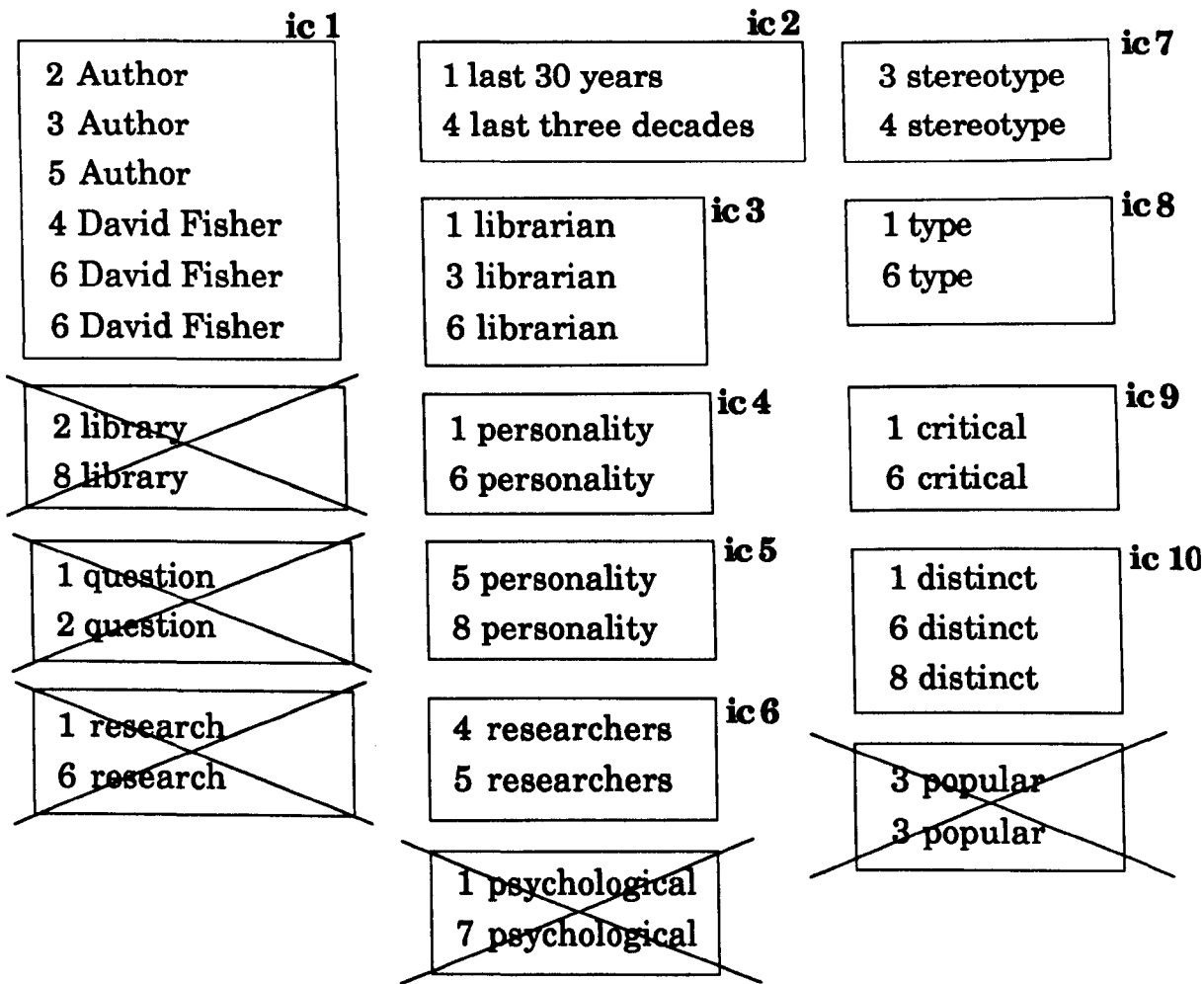


Figure 11.2: Identity Chains in Abstract D

Having identified the IC's, the similarity chains (SC's) can now be considered. Whereas items in IC's are related co-referentially, items in SC's are related to each other either by co-classification or by co-extension (Hasan 1985/1989b: 84). She describes co-classification (1985/1989b: 74) as follows: 'In this type of meaning, the things, processes, or circumstances to which A and B refer belong to an identical class, but each end of the cohesive tie refers to a

distinct member of this class'. Co-extension, on the other hand, relates items 'within the same general field of meaning' (1985/1989b: 74), and is principally realised by the four sense relations (1985/1989b: 79 - 81): synonymy, antonymy, hyponymy and meronymy.

The formation of SC's is in fact the most troublesome stage. Exactly repeated lexical items of the same syntactic class, but which do not properly belong in IC's, are included as SC's; for abstract D, shown in text order, these are:

<i>lexical item</i>	<i>frequency</i>	<i>clause complex number</i>
question	2	1, 2
psychological	2	1, 7
concludes	2	3, 6
popular	2	3

In addition, lexical items which may occur only once, but which enter into co-classificatory or co-extensive relationships with other lexical items, must be identified.

The fifth and sixth candidate SC's involve ellipsis and have strikingly similar grammatical structure. Both "ascertain" (4) and the phrasal verb "points out" (5) are projecting verbs, each one of which has two projected clauses related paratactically. The second ('β+2) merely extends the first ('β1) and has the same projecting verb implicit. It is therefore appropriate that these two SC's are treated as examples of co-classification.

Examples involving relationships of co-extension are most difficult to identify, and it is here where variation between different people's intuitions is most prevalent. Hasan acknowledges the problem of maintaining inter-analyser reliability with regard to collocational ties (1984: 195), so analytical triangulation was carried out to compare and contrast the researcher's own intuitions with those of two other researchers² enlisted to help. This was a valuable exercise, though some disagreements remain. For example, while it seems plausible to suppose that the lexical item "methodologies" of clause complex 1 is synonymous with the lexical item "methods" of clause complex 6 (and should therefore be treated as being co-members of an SC), it is more difficult to account for the intuitions of one of the researchers, who wanted to chart a relation between "addresses" (2) and "reviews" (4).

²Sandra Williams and Dawn Wright, the two other members of BT's automatic text summarisation project team.

An attempt to resolve conflicts of this kind could be made by making appeal to an electronic thesaurus, but it must be remembered that these are notoriously difficult to construct, and often contain inconsistencies, omissions and highly variable degrees of semantic correspondence between items. An analysis of transitivity (Halliday 1985: chapter 5) might be a useful complement to a thesaurus, in that it can show how essentially the same types of process and participants are involved in the "analysing" of 1 and the "discussing" of 5, for example. The part that the lexical items play in the grammar of the clauses can affect the derivation of SC's; for example, in Figure 11.3, both "addresses" and "reviews" feature in α clauses, whereas "discussing", which occurs in an agentless β , is matched with "analysing", which occurs in a downranked clause acting as an Adjunct. Whether or not this type of grammatical information should affect the construction of SC's is not immediately obvious from Hasan's treatment of her simple narratives, and needs further discussion.

For the time being, however, there remains no principled means of deciding whether or not the two instances of the postmodifier "used" (1 and 6), and possibly the "conducted" of 1 also, form a bona fide SC. Perhaps verbs should only be considered candidates for inclusion in SC's if they occur as verbs of main clauses, but until more empirical studies are undertaken to refine the analysis, this must remain speculation.

Another major stumbling block concerns the number of lexical items assumed to form a single link in a chain. Hasan allows only single lexical items as chain constituents, but this constraint disallows complex dependencies which are intuitively felt to obtain between such items as "psychological research" (1), "major studies" (4), "psychology" (2) and "the whole psychological approach" (7).

So, whereas the IC's could be diagrammed with some confidence, the following bundling together of SC's is altogether more conjectural:

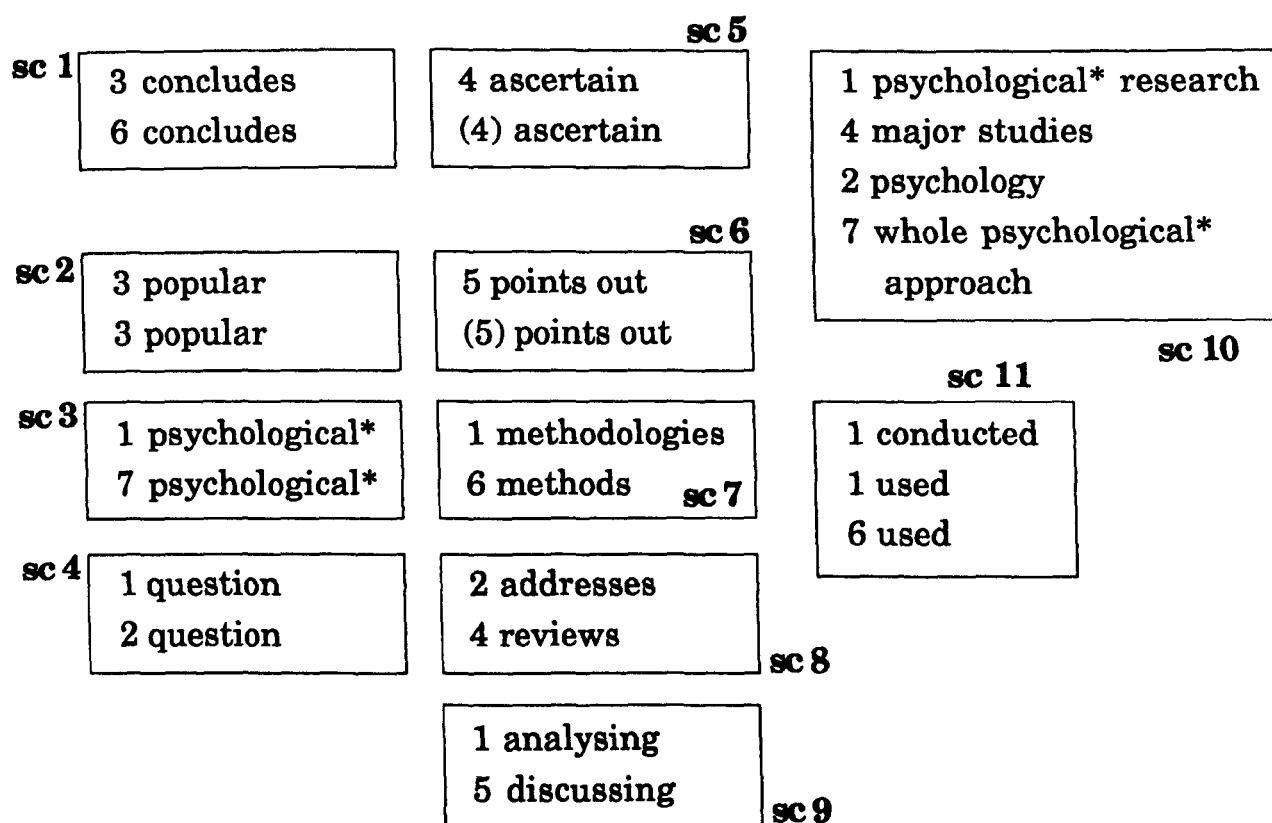


Figure 11.3: Similarity Chains in Abstract D

Brackets enclosing clause numbers indicate an elided element; lexical items flagged with an asterisk indicate the same lexical item occurring in more than one SC. This has been done to show the consequences of allowing more than one lexical item to form a link in an SC; presumably if multiple lexical item SC's are to be allowed, then the decision must be made whether or not to chart the interactions of their component lexical items. Once again this would require further investigation.

11.2.1.3 Plotting Chain Interaction

Hasan's third stage in the analysis of cohesive harmony is the plotting of the interactions between the different chains (1985/1989b: 91 - 94); this shows how similar component messages in the text combine to form a recognisable textural whole. Hasan says that the relations which bind together these similar messages are essentially grammatical and goes on to say (1985/1989b: 91) 'A minimum requirement for chain interaction is that at least two

members of one chain should stand in the same relation to two members of another chain'.

Hasan claims that the following five relations (1985/1989b: 93) hold between interacting chains:

- 1) 'actor action'
- 2) 'action acted-upon'
- 3) 'action and/or actor location'
- 4) 'saying text'
- 5) 'attribute attribuand'

In the abstracts data, computing the chain interactions was slightly less problematic, but performing this operation successfully presupposes that all and only the correct IC's and SC's have already been identified. As has been suggested above, this cannot be guaranteed. Figure 11.4 must then be regarded as a 'best guess' at describing the chain interaction in abstract D.

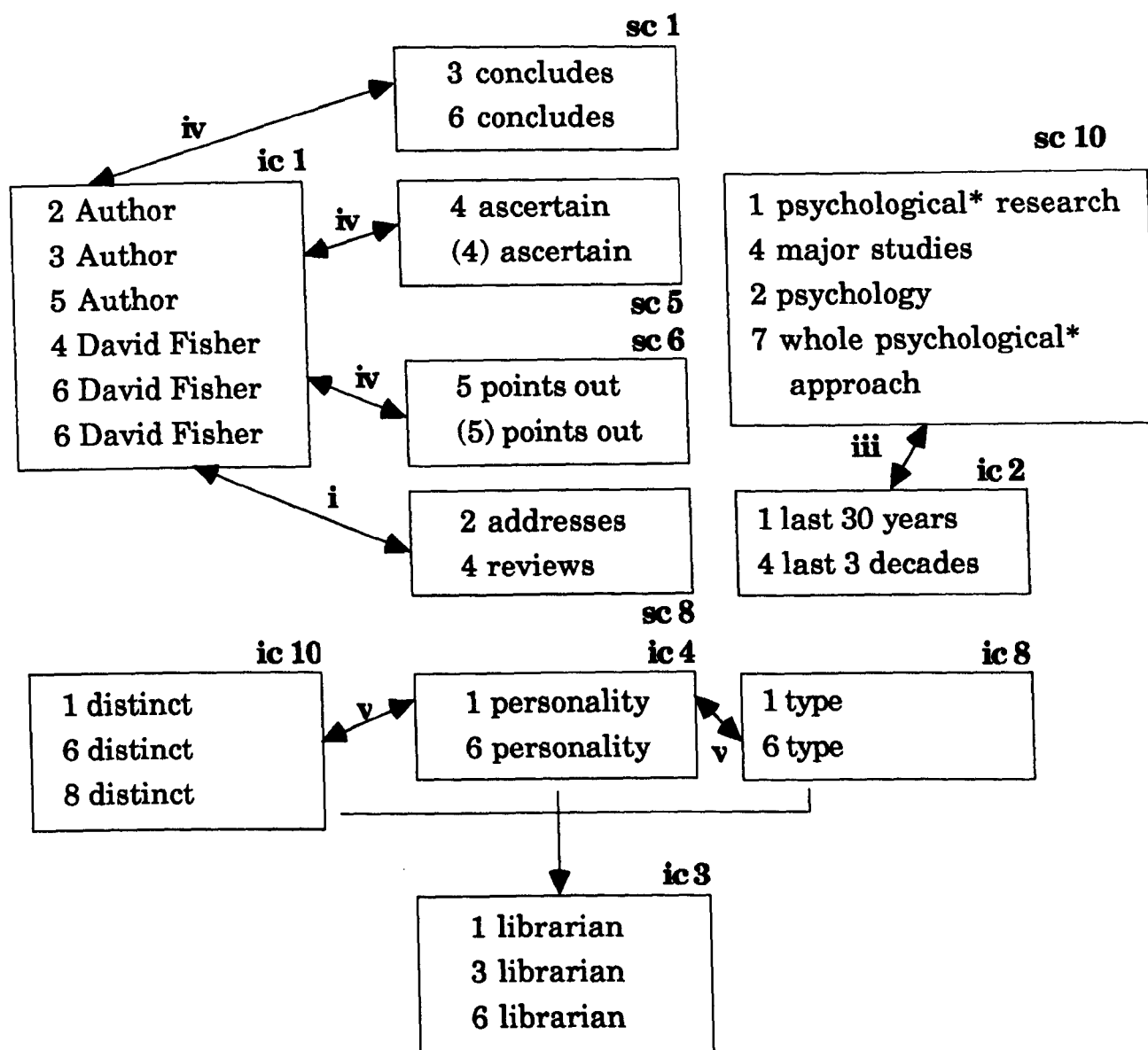


Figure 11.4: Chain Interaction in Abstract D

The analysis of chain interaction reveals a pattern of three separate chains with differing numbers of links.

The first chain, that centring around the first IC, suggests that there is one agent (David Fisher, the author), who, throughout the course of the abstract, performs a number of similar verbal actions as indicated by SC's 1, 5, 6 and 8, three of which involve projected mental processes. These linkage types are presumably analogous to Hasan's fourth type of link, 'saying text' (1985/1989b: 93).

The second chain, concerning the amalgamation of SC's 3 and 10 (the psychology work concept) with IC2 (its temporal span), is more marginal, and

requires that Hasan's third type of linkage be extended to incorporate temporal as well as spatial location. This second cluster serves to indicate that there is a subject area which is regularly talked about in the abstract (SC's 3 and 10), a subject area which is associated with a particular temporal duration on more than one occasion (IC2).

The third chain is perhaps the most problematic, cohesion being engendered by the lexical items "librarian", "personality", "distinct" and "type", four lexical items central to the overall message of the journal article which also occur in its title. It is clear that, depending on the exact definitions of the linkages, these dependencies could be diagrammed in many different ways. One of the problems here is that the framework has no proper means of describing embedding. Thus, the link between "personality" and "type" can be thought of as being epithet - thing (Hasan 1984: 216), but the lexical item "distinct" premodifies both IC4 and IC8. Further the relationship between these three lexical items and the librarian concept, in a sense the thrust of the whole article, cannot be satisfactorily captured by the existing framework. The linkage which binds IC3 to all three lexical items, "distinct", "personality", and "type" (again something which is hard to represent diagrammatically), is probably closest to attribute - attribuand (Hasan 1985/1989b: 93), but is better thought of as being a kind of relational process, something which the existing framework takes little account of.

It is at this point when one would ordinarily undertake the fourth stage in the analysis, the computing of the various ratios. These are different between the Hasan's two articles, but essentially what happens is that each token is put into one or more of four categories: a particular token can be considered 'peripheral' if it does not belong in any chain; on the other hand, a token can be considered 'relevant' if it forms part of an IC or SC; relevant tokens can either be 'central' if they belong in chains which interact with other chains; or 'non-central' if they belong in isolate IC's or SC's. This last stage will not be reported on here, however, since any results which may accrue from the ratios would have to be balanced against the confidence with which the IC's, SC's and their interactions were plotted.

Overall it was felt that this particular framework is not sufficiently formalised to justify its deployment in full-scale analysis. This is not to imply that a perfect framework is sought, but rather that the descriptive text analyst must always keep in mind not only what needs to be done in theory but also what can be done in practice.

The fact that cohesive harmony was not taken further in the present study

should most definitely not be taken to mean that there can be nothing interesting to be said about the patterns of cohesive harmony found in abstracts, however. Intuitively, one feels that there are many interesting findings which would emerge from an analysis of cohesive harmony. Currently, however, the framework is not sufficiently well refined to enable precise hypothesis testing. The reader is referred to Parsons (forthcoming), which constitutes a thesis-length attempt to make the analysis more usable.

Refinements will inevitably have to be made to any linguistic description, especially when applying it to untested types of language, but, since refinement is itself typically an extremely time-consuming activity, care must be taken to ensure that 'refinement' does not escalate to fully blown theory building. For this reason another means of investigating cohesion was pilot tested to see if it would yield interesting results for a small number of abstract texts.

11.3 Static and Dynamic Ties

As stated in the introduction to this chapter, Hartnett's system (1986) rests on the claim that there are two clearly separable types of cohesive resource, one static, the other dynamic. Static ties help to hold attention on a particular topic, while dynamic ties can serve to develop it.

Hartnett (1986: 145) identifies the following resources for constructing static ties: repetition of the same lexical item; demonstratives; third-person pronouns; definite articles; nominal, verbal and clausal substitution; ellipsis; parallel structure; choice of tense; and the various sense relations Hasan claims are involved in co-extension (except hyponymy; see below). Dynamic ties, on the other hand, can be formed using the following resources: temporal conjuncts; lexical superordinates; hyponymy; causal and adversative conjunctions; and comparative and superlative adjectives.

Hartnett formulates a number of specific hypotheses concerning the functioning of these two types of tie, most of which have direct bearing on the concept of success. In effect she provides a means of comparing between different texts, arguing that both the number of instances of the two different types of cohesion, and the ratio exhibited by a particular text, will correlate with a holistic rating of that text's overall coherence.

To test this prediction, two of the Distinct Personality Type abstracts (C and D) were so analysed. Static cohesive devices (1986: 145) include lexical

repetition, the use of demonstratives, third person pronouns, 'definite articles (when they serve to maintain attention on a topic previously introduced)', substitutions and ellipsis, continuative and additive conjunctions, 'synonyms, near-synonyms, antonyms, and collocations', parallel structure and tense. Not all of these categories were searched for: collocations, as Hasan has noted, are troublesome, and there can be no clear differentiation between synonyms and near-synonyms - for these reasons sense-relations were excluded from the analysis; Hartnett's remarks concerning parallel structure and its relation to tense are somewhat obscure, so these too were excluded.

Sense-relations were similarly ignored in the search for dynamic ties. Remaining dynamic indicators charted were: temporal conjuncts; causal and adversative conjunctions; and comparative and superlative forms of adjectives and adverbs.

The computation was performed once again with the aid of the word list, candidate ties being checked such that spurious occurrences could be discarded (the computer finds four instances of "as" in abstract C, but only one of these (in clause complex 8) is a genuine causal conjunction, for example). As can be seen from the annotated appendices, there are problems even with such a seemingly automatic process. Appendix 11.1 lists both the static and dynamic ties in DPT C, and 11.2 those in DPT D. The notation in the Appendix is as follows:

author (3); 2, 3.2, 5.2

This means that there are (3) occurrences of the lexical item "author", one in clause complex 2, one in clause 3.2 and one in clause 5.2 (see Appendix 12.1). A table of numerical findings comparing the two abstracts is derived from the appendices and appears as Figure 11.5:

Static	C	D	Dynamic	C	D
lexical repetition	33	39	temporals	0	0
demonstratives	6	1	causals	1	0
3rd person pronouns	3	0	adversatives	0	0
definite articles	4	9	comps/superl's	1	1
subs/ellipsis	1	3			
continuatives	0	0			
additives	1	0			

Figure 11.5: Static versus Dynamic Cohesion in the two abstracts

It is important to remember that the two abstracts are not the same length - C is 230 words long, and D 192, but even when this imbalance is adjusted for, it should be apparent that there are no startling differences between the two, both in terms of the numbers of each type of device, and the ratio of static : dynamic ties. Because the analysis seems to reveal few differences between the two abstracts (chosen because they were judged to be the best and worst overall of the four), and the fact that only a subset of the framework could be reliably tested, the decision was taken not to make further use of the analysis.

11.4 Conclusions

The following conclusions are to be drawn from this chapter:

- Hasan's cohesive harmony technique was piloted because it was thought that some of the judges' qualitative remarks ('lack of flow', for example) might be explained by a lack of unity of texture (Hasan 1985/1989b: 52).
- However, a number of difficulties were encountered when using the analysis. The main reason for not continuing with Hasan's approach would appear to be that, pending Parsons' work on refining the framework (see Parsons forthcoming), it would appear not be a sufficiently reliable tool for the purposes of precise hypothesis testing.

- A number of problems have been identified with the approach, which should be of use to further researchers who may themselves require to refine the framework. The major problem is that, given the existing framework, one cannot yet reliably decide which lexical items should be grouped together to form identity chains, nor can one reliably decide which lexical items should be grouped together to form similarity chains.

- However, the discussion above does suggest some improvements to the existing framework. In addition to Hasan's claims, the following formal stipulations are suggested to make the process for forming identity chains more explicit and thereby more reliable:

a single identity chain should only contain members of the same syntactic class;

adjectives that modify different headwords should be assumed to be non-coreferential.

- Hartnett's work on static and dynamic ties, on the other hand, with certain modifications, was found to be decidedly more usable overall, it being in essence a much simpler system.

- However, it proved less than useful in helping to differentiate between good and bad abstracts.

12

Analysis of Grammatical Intricacy

12.1 Introduction

This chapter investigates the bearing grammatical intricacy has on the success of a text, as perceived by the reader. Although a relatively new term, grammatical intricacy at first sight would seem to be linked with measures of readability and syntactic maturity, terms which have much longer histories. The chapter starts therefore with a brief discussion of these more familiar concepts, and reviews the various proposals for measuring them.

Halliday's description of grammatical intricacy is reported in section 12.3, and it is argued that its two facets can best be measured by using the framework Halliday proposes for describing the way clauses combine to form larger structural units (section 12.4). This type of analysis, most comprehensively articulated in chapter 7 of Halliday 1985, is exemplified in section 12.4.1. Criticisms that have been made of this part of the theory are noted in 12.4.2.

A first hypothesis is stated in 12.5 and the rest of the section describes how this was tested; two practical problems concerning 'ungrammaticality' (12.5.2.1) and segmentation (12.5.2.2) are also described.

Results for this first hypothesis are then detailed, together with the relevant tables showing the numerical findings. All results reported in this chapter have been derived from the analyses presented in Appendix 12.1, in which all seventeen Information Science abstracts are segmented first into clause complexes and then into clauses.

Further testing is reported in section 12.7; again, practical problems are described, this time concerning dependency structure (12.7.2.1) and the determination of the logico-semantic relations (12.7.2.2). Mainly because there are so many different ways in which clauses may be joined together, the discussion proceeds inductively rather than deductively (see Part Four Introduction): hypotheses relating each of these to the judgements of success

are not formally stated; instead, tables of the distributions of the various linkage types are presented for each set of abstracts (12.8.1 - 12.8.4), and correlations between these and the observed preferences are then discussed. Where appropriate, reference is also made to the judges' qualitative remarks. Section 12.9 reviews the results, and, finally, section 12.10 provides some conclusions.

12.2 Readability and Syntactic Maturity as Determinants of Success

This section is intended to provide a review of the various proposals for measuring readability and syntactic maturity. Although readability and syntactic maturity formulae are used for very different ends - typically readability measures are used to assess the complexity of adult writing, often with a view to deciding what can be given to children to read and when, whereas syntactic maturity formulae are used to measure children's writing developmentally - in practice the two are measured using similar tools and techniques.

Although there are very many different formulae, three common factors emerge: 'First, almost all formulas have some measure of word difficulty. These usually turn out to be a direct or, more commonly, an indirect measure of word frequency. Second, about 60 per cent of the available formulas use some measure of sentence length. Third, about 30 per cent use some measure of sentence complexity' (Pearson 1974: 159).

To return to the distinction made in Chapter 1 above, such formulae are internal, rather than external, measures: typically they are based upon certain readily quantifiable characteristics of the text, and are used to predict how difficult to read it will be. The lack of correlation between these formulae on the one hand and external measures on the other (the perceived level of complexity as reported by the reader, for example) has brought the majority of these techniques into disrepute. However this chapter aims to show that certain measures, those which are relatively hard to derive in fact, can provide some interesting insights into how syntactic complexity might affect perceived success. It is important to point out that, because grammatical intricacy measures linguistic phenomena rather than readers' perceptions or expectations directly, it is still only an *internal* measure, as are the various readability and maturity indices. However, it is suggested that measures based

on grammatical intricacy can *better model* the judges' opinions, opinions that were measured *externally* (see the discussion above in Chapter 8).

The three sections which follow give an overview of the different kinds of formula: firstly, those which make no appeal to grammar (12.2.1); those which chart surface level syntactic phenomena (12.2.2); and finally (12.2.3), those which make psychological claims about processing difficulty by measuring the number of steps required to transform series of simple underlying deep structures into complex surface forms.

Throughout the course of this chapter, various means of measuring syntactic complexity will be derived which concentrate on surface forms, and some of these will make appeal to *depth*, a concept which has received greatest attention in transformational generative theory.

12.2.1 Simple Readability Formulae

Campbell and Holland 1982 and Harber 1979 provide useful overviews of the different readability formulae which do not presuppose any knowledge of syntax. Writing from a predominantly practical teaching viewpoint, Perera 1982 provides arguments from the levels of lexis, grammar and discourse why readability formulae should be treated with the utmost caution. Harber reports (1979: 438) that she 'reviewed 22 readability formulas and found that 18 included estimates of syntactic complexity. Of these 18, 12 use average sentence length in words'. Mean sentence length, however, has been shown to be a poor predictor of syntactic complexity.

Although such indices have little explanatory power, they can be quickly and easily computed, a fact which might explain their continued existence. Nisus®, the word processor used to write this thesis, for example, as well as telling authors how many words they have written, also provides measures of 'Flesch Reading Ease' and 'Reading Grade Level'.

The algorithm for computing the second of these formulae is as follows:

$$\begin{aligned} &0.39 \times (\text{average number of words per sentence}) \\ &+ 11.8 \times (\text{average number of syllables per word}) \\ &\text{Total} - 15.59 = \text{Grade Level} \end{aligned}$$

(Source, Grammatik Mac 1990: Chapter 6, page 10)

Such measures are easily discredited. For example, according to Nisus®, 'This morning I went to the shops. I bought myself some earrings' has a Reading Grade Level of 2, whereas the single sentence 'This morning I went to the shops and I bought myself some earrings' has a Reading Grade Level of 5. This contradicts work by Hunt 1965, who found that the second formulation is not appreciably more complex than the first.

Many of the formulae which purportedly measure syntactic complexity are in fact doing no such thing: this is because the sentence is defined orthographically, rather than syntactically. In contrast, the next two sections review studies which do make genuine reference to syntactic structure.

12.2.2 Formulae Based on Surface Structures

Perhaps one of the best known names connected with this kind of work is that of Kellogg Hunt (see for example 1965). Hunt was one of the first people to recognise the inadequacy of measures based on orthography, and saw the need for a syntactically motivated unit of analysis. His formulation of the T-unit, standing for 'terminable unit' and comprising the main clause together with any subordinate clause or clauses (1971: 294), sparked off countless investigations into the mean number of words per T-unit, mean number of clauses per T-unit, etc. (see Vogel 1985 for details of further studies based upon the T-unit).

Hunt has always maintained that his work was meant to examine syntactic maturity (see for example 1977), and was not intended only as a measure of complexity. His work is of particular importance for this study for two reasons: firstly, many of his studies and those derived from them, including the Indices of Syntactic Maturity (Christensen 1968, Dixon 1972) and the Syntactic Density Score (Golub and Kidder 1974), have attempted to explain why some texts are considered better than others; and secondly, it has led to more fine-grained investigation of syntactic phenomena. The SDS, for example, measured not only the average number of words per T-unit, but also the number of subordinate clauses per T-unit, the mean word length of main clauses, the mean word length of subordinate clauses, the number of modal verbs, the number of be/have auxiliaries, the number of prepositional phrases, etc.

Botel, Dawkins and Granowsky 1973 carried out similarly focused work by checking individual sentences for particular configurations of syntactic features, each ranked 0, 1, 2 or 3, according to hypothesised processing

difficulty, 0 representing minimally difficult structures, 4 maximally difficult structures.

Level 0 features included:

SV, SVO, SVC structures, and simple co-ordination;

Level 1 features included:

SVOC structures, adverbs, modifiers, PPs, and modals;

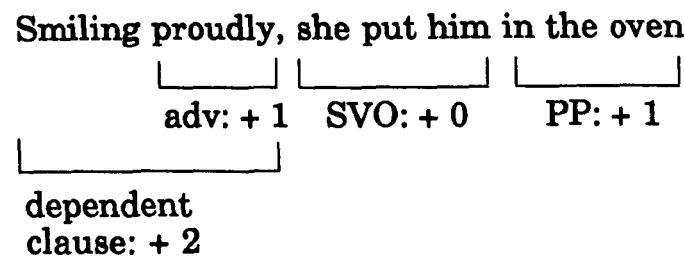
Level 2 features included:

passives, dependent clauses, and comparatives;

Level 3 features included:

clause as Subject.

The following sentence would therefore receive a score of 4 as follows:



The complexity score for a text was a function of its component sentences: 'The syntactic complexity of any passage or sampling of sentences is the arithmetical average of the complexity counts of the sentences evaluated' (Granowsky and Botel 1974: 34).

More recently, Arena 1982 has assessed the complexity of legal writing primarily in terms of the way clauses combine to form larger units. Though he does not explicitly say so, these larger units, or 'information blocks', appear to be equivalent to Hunt's T-unit (as are Fichtner's 'infrastructures'; see Fichtner 1980). He posits the idea of an information block so 'that both sentence length and the number of conjoinings can be excluded from complexity measures', stating 'Complexity then depends only on the number and kinds of clause embeddings' (1982: 146).

The analysis he advocates proceeds in two stages: 'First, the written text is literally broken down into single clauses, rewritten one clause to a line, and numbered consecutively ... Secondly, each clause is labeled with a letter; and information blocks containing a main clause are separated from each other by boundary symbols' (1982: 146). The boundary symbols used are '#', to delimit sentences, and '+' to delimit information blocks. Arena demonstrates this

technique by providing an analysis of nine sentences taken from a 'legal opinion by a Du Pont attorney'. The second of these nine sentences is as follows (the first contained three clauses, so the numbering starts at four):

- # A 4. The part of the Examiner's response (5) (6) suggests (7)
- B 5. that is critical of this showing
- C 6. because of the alleged absence of a property difference
between precrimped and non-crimped filament and fabric
- D 7. that Applicants have failed
- E 8. to make it clear
- F 9. that the property (10) is crystalline memory force (11)
- G 10. relevant to precrimped filament
- H 11. and (is) not shrinkage force. #

(taken from Table 1, 1982: 147).

According to Arena, then, this is one orthographic sentence, consisting of one information block, or T-unit, which comprises 11 clauses. There are a number of things to be noted here. Firstly, although Arena does not explicitly provide a definition for the clause, it would appear that the presence of a lexical verb is taken to be a strong indication (although this would not explain why he has chosen to make 6 a separate clause). Secondly, in keeping with transformational practice, no distinction is made between what Halliday would call embedded clauses (for example, clauses 5 and 6) and dependent clauses. Thirdly, concerning the notion of subordination more generally, Arena explains the symbols he uses (1982: 147) as follows:

- A = a main clause
- B = a subordinate clause embedded into A
- C = a subordinate clause embedded into B
- D = a subordinate clause embedded into C, etc.

Therefore, in the above example, clause 7 is subordinate to clause 6, since it is represented by the letter D, which is subordinate to clause C. This is rather a curious claim, as clause 7 immediately follows clause 4 in the surface structure, and tells the reader what is being suggested.

Arena proposes three complexity measures (page 149) - average number of clauses per sentence, average number of clauses per information block, and

embedding depth - and concludes (page 153) that the second and third are most useful. The third of these measures adds to what has been discussed so far, and foreshadows some of the discussion in the next section. Briefly, Arena claims that embedded material must be held longer in memory than non-embedded material, so that information blocks with many clauses will be perceived to be more complex than those with few clauses. This is an interesting and important observation which will be investigated later in the chapter.

Arena also provides a measure of 'embedding depth' based on the claim that 'the H clause takes eight times as long to process as the main clause' (1982: 151). The exact means of calculating the depth score is not important, although it should be pointed out that it crucially depends on the rather idiosyncratic view of dependency mentioned above.

The idea of 'depth' is an important one for the purposes of this chapter because it was hypothesised that some of the informants' objections to verbosity and the like could be explained by making appeal to a certain kind of depth score. Some of these judges' criticisms are as follows:

Tanzania D:

Suffers from omissions ... and verbosity (judge 8).

Tanzania E:

verbosity detracted from E's appeal (judge 8).

Distinct Personality Type C:

poor grammar and punctuation. Too conversational (judge 4).

Vickery and Vickery, Chapter 1 B:

B is not easily read, but its individual sentences are clear and well-constructed (judge 8).

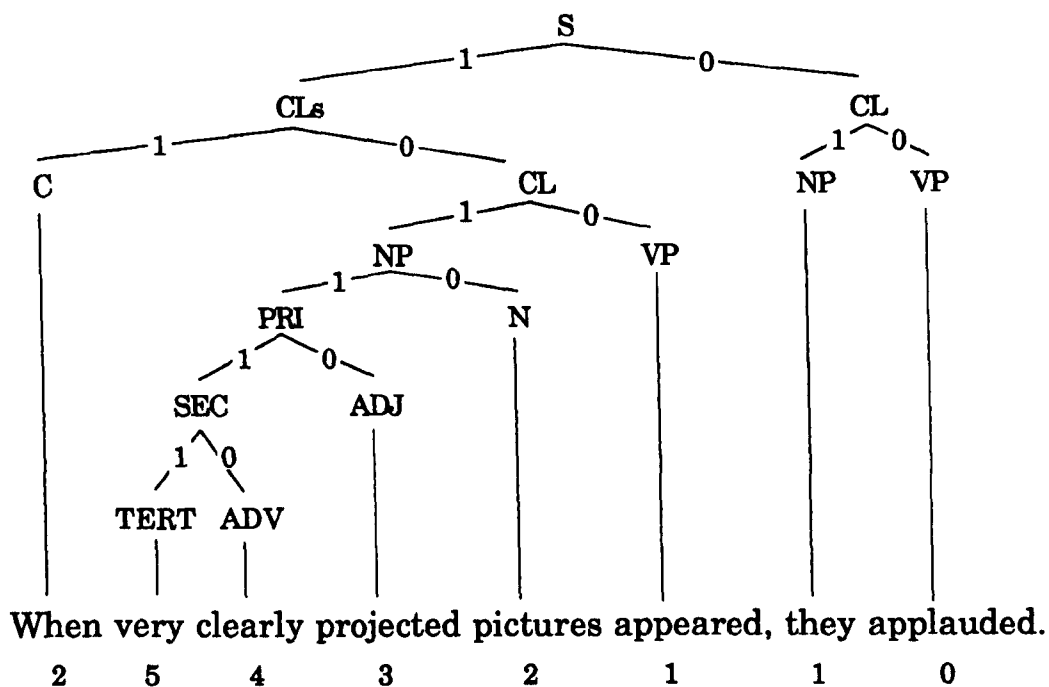
The next section situates the depth concept in more detail by relating it to what is called the 'Derivational Theory of Complexity'. DTC has provided the impetus for the investigation of depth discussed later in the chapter, although in this work depth has been measured in a very different way, as will be seen below.

12.2.3 Formulae Based on Derivational Complexity

The basis of the Derivational Theory of Complexity (DTC), generally accredited to Fodor and Garrett 1967, is that 'sentences are more complex if their derivation from base-to-surface structure, in a transformational grammar, involves a greater number of transformations' (Schluroff 1982: 135).

Progress into DTC prompted Hunt to consider the relationship between the two levels, and by conducting a series of experiments in which children were asked to rewrite short simple sentences to make them longer but without omitting anything important, he was able to hypothesise various deep structures and the transformations required to derive the surface structures. Hunt showed that syntactically more mature children were better able to consolidate a large set of simple sentences into one complex sentence, concluding (1971: 290) 'as schoolchildren get older they tend to embed a larger and larger number of sentences within some uppermost S constituent'.

There are many other psychological process models referred to in the literature on readability, but one which has particularly influenced the measures which are derived in this chapter is generally known as Yngve's depth hypothesis (Yngve 1960). Yngve proposed a means of assigning a depth score to a sentence, which was taken to represent the amount of temporary storage required to generate it. The algorithm for computing the sentence depth was as follows: 'First, number the branches of each node from 0 to $n - 1$, where n is the number of branches from that node. Start numbering from the right ... Then, compute the depth d of each terminal node by adding together the numbers written along all branches leading to that terminal node' (Yngve 1960: 450).



In the above example, therefore, the largest value of *d* is 5, and is taken to be the depth of the sentence. It should be noted here that Yngve depth makes claims about the relative amounts of processing required for right- and for left-branching sentences. Heavily right-branching sentences are considered easier to process than heavily-left branching sentences: this is because the numbering of the branches proceeds right to left.

As will be seen later in the chapter, a similar kind of depth measure is proposed for the quantification of grammatical intricacy. Although influenced by the types of model discussed in this section, it is different in three ways: firstly, no consideration is given to deep structure; secondly, measures are based on the way whole clauses combine, rather than on the way sentence constituents combine; and thirdly, there is no bias between left- and right-branching constructions: each receive equal weight. This will be returned to below in section 12.5, after considering the role of grammatical intricacy as a predictor of perceived textual complexity.

12.3 Halliday's Conception of Grammatical Intricacy

As Halliday has recently pointed out (1988: 176), complexity cannot be accounted for by any single linguistic notion. It is a composite of at least two different dimensions: the lexical and the grammatical. Lexical density was

investigated in Chapter 9 above; this chapter will focus on grammatical intricacy.

Although grammatical intricacy (Halliday 1989: 76 - 91) is very rarely discussed without reference being made to lexical density, the view taken in this chapter is that the two are very different types of phenomenon. Halliday explains their interrelationship in terms of the cline between speech and writing, noting that writing tends to be lexically dense and grammatically simple, whereas speech tends to be lexically sparse and grammatically intricate. Arguably, the concept of lexical density is better understood, and certainly is more easily measured: one of the reasons why the present chapter is more complicated than the chapter on lexical density above is that there does not yet exist an accepted standard for quantifying grammatical intricacy.

Although Halliday has so far been reluctant to furnish an explicit definition, grammatical intricacy seems to pertain to the extent to which clauses combine with other clauses to form larger units. These larger units are termed 'clause complexes' and 'are the grammar's way of showing (1) *that* and (2) *how* the processes going together in a sequence are all related to each other' (Halliday 1989: 82, original emphasis). Halliday's concern is to show that, contrary to what is commonly believed in language studies, writing is typically less grammatically complex than speech.

This thesis is about abstracts and abstracting, and is therefore not particularly concerned with comparisons between speech and writing. However, it is the contention of this chapter that the means by which written texts are analysed firstly into clause complexes and then into their constituent clauses (most comprehensively articulated in Chapter 7 of Halliday 1985) can help to explain why some are perceived to be more grammatically complex than others. Clause complex analysis will be reviewed in the next section.

12.4 Clause Combining as a Determinant of Grammatical Intricacy

It is important to stress that it is not the assumption of this chapter that 'complexity' can be wholly accounted for in terms of the type of analysis Halliday proposes in Chapter 7 of IFG. Rather, just as it is assumed that success is a multivariate notion only partly determined by the reader's perceptions of complexity, grammatical intricacy is assumed to be a

multivariate notion only partly determined by complexity at the clause level. As the review of the readability and syntactic maturity literature in section 12.2 above demonstrated, other researchers have proposed measures of complexity at different linguistic levels and in altogether different ways.

Why then has the decision been taken to adopt the Hallidayan approach to clause combining as the preferred means of analysis? The answer is in two parts. Firstly, although many commentators have often been very stern in their objections to certain facets of systemic-functional grammar, they nevertheless concede that it provides a workable framework particularly well suited to the description of *real texts* (see for example Hudson 1986 and Huddleston 1988). Secondly, although there is much in Chapter 7 of Halliday 1985 which covers old ground, and in many ways is not so very different from transformational grammar or from more traditional grammars, there is also much that is new, and of a distinctively systemic flavour. One of the contentions of this chapter, for example, is that monitoring the 'logico-semantic relations' between clauses (concepts which are uniquely Hallidayan) can provide useful insights into the nature of the success judgements, insights which would not necessarily have emerged from other kinds of analysis.

Halliday has implied that grammatical intricacy is largely a function of 1) the extent to which clauses combine to form larger structural units, and 2) the way in which these combinations are realised, and so two sets of hypotheses are investigated in this chapter: the first hypothesis (12.5) examines the first of these assumptions - that is, the degree to which the *amount* of clause level complexity can predict informants' preferences; the second set (12.7) examines the second - that is, the degree to which the *types* of combination can predict their preferences.

12.4.1 Clause Complex Analysis

Prior to hypothesis testing, however, it is proper that the idea of 'clause complex analysis' is discussed a little more fully. Analyses of theme, transitivity, cohesion, and so on, are widespread, and are acknowledged means of describing data. The status of clause complex analysis, on the other hand, is less recognised, and, although several researchers have used Halliday's insights into clause combining to advantage (see for example Drury 1989a, Martin 1986b, and Butt 1984), they are altogether less well known.

With this in mind, a small section of the analysis will be described in some

detail in the main body of the text. The first half (six of the thirteen clause complexes) of abstract version D written for the Tanzania article will be examined below; this particular sample was selected because it has a large ratio of clauses to clause complexes, and nicely exemplifies many of the phenomena Halliday documents in his seventh chapter. Problems with this type of approach will be treated in sections 12.4.2, 12.5.2 and 12.7.2 below.

The first two clause complexes each contain only one clause:

1

The article is about the development of library and documentation services in Tanzania.

2

The paper examines the needs against the source resources and against the unlimited demand.

The third clause complex is a little more complicated:

3 (1¹ ^ +2)

3.1

Due to financial, political and moral support the library service has taken 20 yrs□ to reach 1% of the population,□

3.2

and the paper puts forward different strategies for a better library service.

The notation for this clause complex - (1¹ ^ +2) - will now be explained in some detail.

Clause complex 3 consists of two clauses, labelled 3.1 and 3.2. These are independent, paratactically related clauses, and are represented using numerical notation (Halliday 1985: 195). The prefix before the second clause indicates the type of logico-semantic relationship that exists between the two clauses; in this case '+' shows that the second clause is an extension of the first, since it adds 'some new element' (1985: 197). In this way, each particular combination of clauses makes a choice a) between the two -tactic alternatives (i.e. hypotactic and paratactic), and b) between the five different kinds of logico-semantic relation (i.e. elaboration (=), extension (+), enhancement (x) [these three being particular kinds of expansion], locution (") and idea (') [these two

being particular kinds of projection]). The clause complex structure is therefore shown as $(1 \wedge +2)$, following Halliday's own notation in Chapter 7 of IFG.

However, this chapter makes use of further notation in the form of superscripts. This is because it is important to notice that clause complex 3 contains some rankshifted, or embedded, material. For reasons that will be gone into below, downranked clauses are shown; these are notated in the usual way using double square brackets. A superscript attaching to a clause number is used to denote the number of embedded clauses contained in that clause; for example, $(1^1 \wedge +2\alpha \wedge +2''\beta^1)$ would signify a clause complex consisting of three clauses, the first and third of which each contain one embedded clause.

Clause complex 4 is altogether more complicated, and consists of five clauses:

4 $(1\alpha \wedge 1''\beta \wedge +2 1 \wedge +2 +2\alpha \wedge +2 +2''\beta)$
4.1
It says
4.2
the library is a precondition in an underdeveloped
country, rather than a luxury
4.3
and puts forward sub-concepts,
4.4
but states
4.5
that options should be kept open.

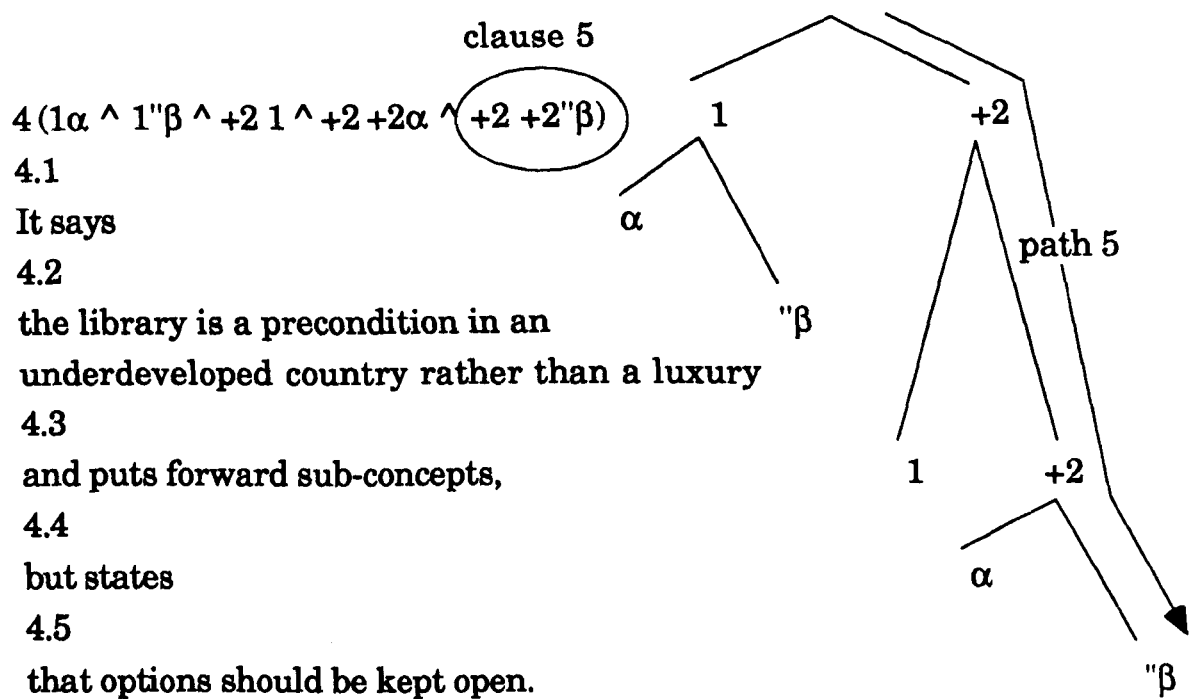
The chosen way of representing the clause complex structure (shown in brackets after the integer representing the clause complex number) is in fact an amalgamation of the two types of notation Halliday provides on page 201 of IFG. These are:

- 1) $\alpha \wedge \beta\beta 1 \wedge \beta\beta 2\alpha \wedge \beta\beta 2\beta 1 \wedge \beta\beta 2\beta 2 \wedge \beta\alpha\beta \wedge \beta\alpha\alpha$
and
- 2) $\alpha \wedge ''\beta (x\beta (1 \wedge +2 (\alpha \wedge ''\beta (1 \wedge +2))) \wedge \alpha (x\beta \wedge \alpha))$

Although this is really only a matter of personal preference, this chapter

prefers to show *constituency* using repeated symbols (as in notation 1), rather than using bracketing (as in notation 2). Repeated symbols are used to designate the 'pathname', as it were, of a clause, making it easier to see its position in the clause complex (see the diagram below). The use of repeated symbols constitutes what Halliday terms 'represent[ation] as at the foot of the tree' (1985: 201). The Greek letters show *dependency*; and the prefixes show the type of *logico-semantic relation*.

Each clause is analysed as a combination of symbols denoting its place and function in the overall clause complex. So, for example, the fifth clause is denoted +2 +2"β. It may help to visualise this as a path from the root of the tree down to the clause leaf:



The diagram above shows the five paths which constitute the tree. The tree is a slightly unusual shape; in the interests of clarity, the branches have been arranged so that the clauses line up with their relevant nodes.

Clause complex 4 contains two projections which are realised as locutions, or constructions of wording ("says" and "states"), as opposed to meaning. Again, the expansions are both extensions, because they add new elements (this particular clause complex is returned to in the discussion of problems, see section 12.7.2.1 below).

Clause complex 5 is analysed as follows:

5 (1 ^ +2 1α ^ +2 1"β ^ +2 +2α ^ +2 +2"β)

5.1

Tanzania has tried to create
model libraries

5.2

but the paper says

5.3

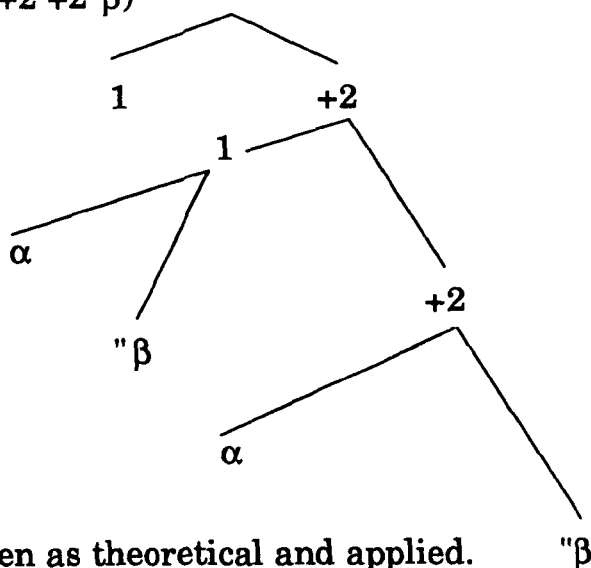
this view lacks perspective,

5.4

and says

5.5

that librarianship should be seen as theoretical and applied.



Again, this sentence is a complex of locutions and extensions. Arguably clause 5.1 contains an embedded clause ("to create ..."), but, following Halliday's addendum to Chapter 7, "tried to create ..." is treated as a hypotactically extended verbal group, the link being one of conation (1985: 258). In other words, it is neither a downranked clause, nor a separate full clause. Although this is 'complexity', it is complexity at a lower level, and so is not explicitly shown in the analysis.

Clause complex 6 is:

6 (1 ^ =2)

6.1

Some principles, generalisations and assumptions, need to be
"liquidated"

6.2

such as libraries are for the middle and upper classes.

This is a relatively simple clause complex, in that there is one relation linking two clauses: 6.2 elaborates (=) the first by providing a specific example (1985: 196). Once again, "need to be ..." is thought to be an example of group level complexity, and is therefore not analysed.

From this small amount of analysis, it can be appreciated how difficult and time-consuming the clause complex task is. There are also a large number of associated problems, some of which are dealt with below.

The next section reviews the criticisms that have been made of the framework.

12.4.2 Criticisms of Clause Complex Analysis

In this section, firstly criticisms concerning the theoretical adequacy, and in some cases the usability, of clause complex analysis will be discussed. Later, some positive points are noted.

Clause complex analysis has been attacked on many grounds, the most serious of which are a) the distinction between constituency and dependency, and b) the distinction between rankshift and hypotaxis.

Hudson (1986: 805 - 809) makes a number of interesting comments concerning what he believes to be a fundamental confusion in IFG between constituency and dependency. Addressing these criticisms is far beyond the scope of this chapter, however: Hudson himself notes that adoption of his own suggestion - dispensing entirely with the notion of constituent structure and letting dependency do all the work, as it were - would require extensive revision of Systemic theory. However it is proper to point out that the criticisms Hudson makes of IFG apply equally to this chapter.

Huddleston recasts one aspect of this confusion in the form of a discussion of the distinction between minimal and maximal bracketing, which, of particular relevance to the present chapter, he exemplifies in terms of clause complex relations. Basically Huddleston's thesis (1988: 148 - 151) is that it is remarkably hard to draw trees for complex sentences in both a valid and reliable way. Specifically, he argues that Halliday overfavours the $(\alpha \wedge \beta \wedge \chi)$ formulation, when the more maximal $((\alpha \wedge \beta) \wedge \alpha \wedge \beta)$, or possibly $(\alpha \wedge \alpha \wedge \beta \wedge \beta)$ formulation would be more consistent with the rest of the theory. However, Huddleston does concede that a 'flattened out grammar' may indeed be more practical for analysing real texts. It is appropriate to declare this controversy here, since the second set of hypotheses investigated in this chapter refer to depth. Hence it must be kept in mind that there are researchers who would disagree with many of the analyses for the data presented in this chapter (see also the discussion in section 12.7.2.1 below).

Turning now to problems of embedding, Berry (1975: 136) notes that it is often difficult to decide between hypotactic univariate structures and structures involving embedding. Also Huddleston (1988: 144 - 148) has pointed out a number of inconsistencies. In the following two examples,

- a) he left before the vote was taken, and
- b) he left before the debate,

he notes that the obvious functional similarity of these two complexes is obscured by insisting that a) consists of two separate clauses.

Huddleston's various arguments are convincing, and so it is right that such difficulties are borne in mind. The view taken in this chapter is that it is indeed difficult to differentiate reliably between rankshifted items and hypotaxis, and so it was decided that the numerical analysis should make no real differentiation between the two (see below).

This concludes the brief discussion of published criticisms of the clause complex framework. Other problems which were encountered throughout the process of analysing the abstracts are discussed below in sections 12.5.2 and 12.7.2.

To redress the balance a little in the face of all these problems, it should be noted that Quirk et al's grammar is consonant with the way dependency is handled in chapter 7 of IFG, although sometimes the emphasis is slightly different: subordination, for example, is thought to be a particular type of embedding (Quirk et al 1985: 44). Other more novel ideas in chapter 7 of IFG, such as the discussion of the logico-semantic relations, have influenced other works: for example, expansion and projection are each given chapter-length treatment in the Collins Cobuild English Grammar (Collins 1990). Also taking Halliday's chapter as their departure point, Nesbitt and Plum 1988 have conducted some interesting research concerning the distribution of expansion and projection in English, probabilities being assigned to the various choices with respect to frequency of occurrence in a corpus of some 123 texts. One of the things that they were able to show was that the percentage breakdown of expanded clauses taken from the corpus showed a marked preference for extension (the option of elaboration being realised 23% of the time, that of extension 51%, and that of enhancement 27%).

The argument put forward in this chapter, therefore, is that although there are very many serious problems with the theory and interpretation of clause complex analysis, these may be outweighed by the many potentially fruitful discoveries to be made using this particular framework.

The next section proposes a first hypothesis which states that the total number of relations between clauses in the clause complexes of the abstracts may, when suitably normalised, help predict those informants' judgements which concern complexity.

12.5 Statement of Initial Hypothesis

In keeping with the assumptions of the various syntactic maturity formulae, it will be hypothesised that higher levels of clause complexity will find favour with the judges. The first hypothesis can therefore be stated as follows:

- **Hypothesis 1:**

Abstracts with a larger amount of clause level complexity will be generally preferred over abstracts with a smaller amount of clause level complexity.

12.5.1 Testing the Initial Hypothesis

Each of the seventeen abstracts has been analysed, firstly into clause complexes, and then into clauses, with clause level embedding shown in the usual way: the seventeen analyses are presented in Appendix 12.1.

Hypothesis 1 is perhaps the most robust of those investigated in this chapter, because it is immune to the theoretical problems alluded to above in section 12.4.2; that is to say, the tensions between dependency and constituency on the one hand, and between hypotaxis and rankshift on the other, need not necessarily affect this particular hypothesis. The dependency/constituency issue can be controlled for simply by calculating the average number of clauses per clause complex. The hypotaxis/rankshift issue can be controlled for by including the number of clause level embeddings into the measure of the number of clauses. The formula used to measure the clause level complexity of each abstract is therefore given by:

$$1 - \frac{\text{Total number of clause complexes in the abstract}}{\text{Total number of clauses} + \text{total number of embedded clauses}}$$

When this figure is multiplied by 100, the result can be taken to be a percentage score for clause complexity, itself a component of grammatical intricacy. Some examples may help demonstrate the nature of this clause complexity score:

- To have a minimum clause complex score (i.e. 0%), an abstract would have to contain a) n clause complexes, each consisting of not more than one clause, and b) no embedding.
- To have a maximum clause complex score (i.e. 100%), an abstract would have to contain n clause complexes, each consisting of a) a very large¹ number of clauses, or b) a very large number of embedded clauses, or c) both.
- An abstract of mid-range complexity, say with 15 clause complexes, consisting of 22 clauses and 13 embedded clauses, would score $((1 - (15 + (22 + 13))) \times 100) = 57.1\%$.

Clause complexity scores for the Information Science abstracts range from 22.2% up to 72.4% (see section 12.6 below).

Although this calculation has been derived to eliminate the two theoretical problems of dependency/constituency and hypotaxis/rankshift, there are problems of a more practical nature which are altogether more troublesome; these are discussed in the next section.

12.5.2 Problems of Analysis

Unfortunately, as is often the case in linguistics, problems are encountered when using a particular framework to describe real data. The clause complex analysis adopted here is no exception, and while only those problems which have a direct bearing on the data are considered, still the discussion stretches to a number of pages.

To break it up into more manageable proportions, this section discusses only those problems which relate to the first hypothesis. Problems encountered while investigating the second set of hypotheses are reviewed in section 12.7.2 below. In all cases a small number of examples are provided to accompany each problem.

¹It is theoretically impossible for an abstract to score a perfect 100%, because 'very large' would have to mean infinite. For an abstract containing 10 clause complexes to score 99%, the total number of clauses and embedded clauses would have to be 1000 $((1 - (10 + 1000)) \times 100)$.

12.5.2.1 'Ungrammaticality'

An important distinguishing characteristic of naturally occurring data is that typically it will contain many features which cannot be satisfactorily described by a particular analytical model. In the abstracts data, ill-formed constructions are prevalent, and, unlike deviant spellings, which could easily be regularized, the analyst sometimes could not recover what was intended by the original writer.

When grammatical errors occur within a single clause, often there are few problems, since the different recoverable interpretations do not alter the basic clause complex structure. In the example below, it seems likely that one or two function words have been omitted linking the two halves of the complex adverbial group, and so the segment can be designated a single clause with a relative amount of confidence.

Abstract Set: Vickery and Vickery, Chapter 1
Abstract: C
Clause Complex: 1

1
The emergence of information science can be identified with the communication of information the changes of society.

Far more common in the corpus however are those syntactic anomalies which can affect the placement of clause boundaries.

Abstract Set: Tanzania
Abstract: A
Clause Complex: 10

10 ($\alpha \wedge \text{"}\beta\alpha\alpha^1 \wedge \text{"}\beta\alpha=\beta \wedge \text{"}\beta x\beta$)
10.1
The author concludes
10.2
that fundamental librarianship principles [that fit into the context of a developing country] must be fulfilled as cost effectively as possible,
10.3
incorporating cheap manpower,
10.4
and to act effectively within industry and commerce.

In the above example it is supposed that most readers would be confused by the 'and' conjunction introducing the infinitival fourth clause. Presumably the

writer has felt obliged to include the conjunction in an attempt to signal that the fourth clause is still part of what is being projected by clause 10.1. But as a result of this, and the fact that 10.2 contains downranked material, and the fact that 10.3 is something of a dangling participle, it is no longer clear precisely what 10.4 is enhancing. Reversing the order of the projected clauses makes for a reading that is a little clearer (' ... concludes that, to act effectively within industry and commerce, fundamental librarianship principles ... '), though it is still not apparent where best to attach the 'incorporating cheap manpower' participle.

There are many instances of confusing grammar in the data, some of which are flagged '?ungrammatical' in Appendix 12.1. It is not intended that the analysis which appears alongside such examples be regarded as the only legitimate interpretation, but rather that, out of all the candidate analyses, the one provided seems to be the most plausible.

Perhaps not unsurprisingly the difficulty facing the text analyst is knowing where to draw the line between what is reasonable to expect of a descriptive framework, and what is not. A large proportion of one particular abstract, version A of the Distinct Personality Type set, is written in what one judge describes as 'idiosyncratic note form'. The following is an excerpt:

Abstract Set: Distinct Personality Type

Abstract: A

DOUGLASS - tests on potential librarians. results - people orderly, but not compulsive, conscientious, slightly submissive, not anxious, less self-confident.

The MORRISON STUDY - results - Librarians more sure of themselves than 'normal people', conflicts with Bryan.

Discussion of the merits of the various tests - McDermott, Black, Douglas. Douglas and the Minnesota Multiphasic Personality Inventory - same problems as Bryan.

An attempt has been made to analyse this abstract and appears in Appendix 12.1. The fact that the judges unanimously considered this abstract worst overall (the only instance of complete agreement, in fact) is perhaps linked to the difficulty with which this analysis was performed.

Version A of the Vickery and Vickery, Chapter 1 set is similarly deviant in that it incorporates numbered headings and lists:

Abstract Set: Vickery and Vickery, Chapter 1

Abstract: A

4 ?: *no main verb*

2) Factors creating information demand

- need for administrative information
- the growth of commerce and the need to identify potential markets and sources of supply
- educational needs of teachers, students and administrators, brought about by ability to read and write
- literacy

5¹

3) The study of information transfer

Information transfer concerns the processes Π involved in the transfer of information from sources to users Π .

Examples such as this prompted the decision not to provide an analysis for any clause complex lacking a main verb (this decision is also carried over into the next chapter on Theme, see below). In the Appendix, clause complexes missing a main verb are signalled as such, and no structural interpretation is given. This means that 'minor' clause complexes are not segmented into clauses, nor is clausal embedding shown.

With reference to the formula given above, this would mean that clause complex 4 above would be treated as consisting of one clause only with no embedding. Contrast clause complex 5 above, which would be treated as consisting of one clause only, but with one embedded clause.

Finally, there are a small number of abstracts whose authors ran out of time in the class exercises, which in some cases caused their last few clauses to appear ungrammatical; these are flagged '< unfinished >', and are analysed as far as possible.

12.5.2.2 Segmentation

Dividing the abstracts into clause-complexes is carried out according to the orthographic sentence boundaries found in the texts. Dividing the clause-complexes into clauses can normally be performed quite reliably with few problems.

Where difficulties do arise, these most usually concern 'which' constructions whose domain is unclear. In:

Abstract Set: Tanzania
Abstract: E
Clause Complex: 4

4

Therefore it is necessary to plan for short, medium and long term projects which may prevent future chaos or problems.

there are two possibilities: 'which ... ' could be treated as a separate (non-defining) relative clause hypotactically elaborating on the whole of the previous clause, suggesting that it is the entire act of planning the various projects which can help avoid disaster; alternatively, the 'which ... ' can be analysed as a downranked qualifier, showing that it is a particular subset of projects that must be planned for. This is normally a difference in punctuation in written English (where a comma is used to separate the elaborating clause), but the authors of the abstracts are not always consistent in their use of this device.

Another uncertainty encountered in the data is found in the following example:

Abstract Set: Vickery and Vickery, Chapter 1
Abstract: D
Clause Complex: 2

2 ($\alpha \wedge \beta^2$)

2.1

The author suggests

2.2

that information transfer has developed in an urban environment for several reasons, namely the large number of people \square involved in such a social organisation \square , the high population concentration leads to a high density of innovators \square producing original information \square , and the need for a level of control and coordination over the population as a whole.

Here one could argue that 2.2 needs further subdivision, the words 'the high population concentration leads to ... ' constituting a separate clause acting as an exemplification (a type of paratactic elaboration, 1985: 203). This analysis has been resisted, however, since the nominal groups either side have similar function. Indeed the sentence reads better if the clausal element in question is reduced to a downranked qualifier 'the fact [[that the high population ...]]', as the syntactic symmetry of the three named reasons is then preserved.

These kinds of deliberation are largely academic in terms of the first hypothesis, and so testing can proceed with a fair amount of confidence.

Issues of this kind will be returned to below, however, when they will be seen to have a good deal more consequence.

12.6 Results of Testing the Initial Hypothesis

The following table of calculations has been derived from the analysis of the seventeen abstracts presented in Appendix 12.1:

Clause Level Complexity in the Seventeen Abstracts

	Total Number of Clause Complexes	Total Number of Clauses	Total Number of Embedded Clauses	Total Number of Clauses + Embedded Clauses	Clause Complexity Score (%)
Tanzania A:	11	19	8	27	59.3
Tanzania B:	4	7	7	14	71.4
Tanzania C:	8	11	9	20	60
Tanzania D:	13	31	4	35	62.9
Tanzania E:	13	22	17	39	66.7
Distinct A:	43	59	11	70	38.6
Distinct B:	12	27	14	41	70.7
Distinct C:	13	18	13	31	58.1
Distinct D:	8	19	10	29	72.4
Distinct E:	6	9	6	15	60
V & V, Chapter 1 A:	13	15	5	20	35
V & V, Chapter 1 B:	13	27	10	37	64.9
V & V, Chapter 1 C:	14	15	3	18	22.2
V & V, Chapter 1 D:	11	19	6	25	56
V & V, Chapter 2 A:	13	16	10	26	50
V & V, Chapter 2 B:	16	24	4	28	42.9
V & V, Chapter 2 C:	7	15	3	18	61.1

The first hypothesis stated that abstracts with a larger amount of clause level complexity would be generally preferred over abstracts with a smaller amount of clause level complexity.

In the case of the Tanzanian abstracts, judges generally preferred version E (scoring 34 points), then version C (scoring 26), with little to separate B, A and D (scoring 21, 20 and 19 respectively). This will be referred to as the observed rank. The rank predicted by the hypothesis using the 'Clause Complexity Score (%)' measure is B, E, D, C, A. This will be referred to as the expected rank.

Abstract Version	A	B	C	D	E
Preference Score	20	21	26	19	34
Complexity Score (%)	59.3	71.4	60	62.9	66.7
Observed Rank	E	C	B	A	D
Expected Rank	B	E	D	C	A

There seems to be only a weak correlation between these two ranks; abstracts in the observed rank are never more than two positions away from abstracts in the expected rank. However, overall, the hypothesis is not supported for the Tanzanian data.

This is somewhat surprising, since many of the judges' qualitative opinions concerning the Tanzanian abstracts would seem to suggest issues of readability and quality of writing. A selection of their comments are reproduced here:

- A:
quality of the writing not too good (judge 2);
too disjointed. I prefer proper sentences (judge 3);
incoherent (judge 4);
very jerky portrayal of concepts, Didn't flow at all (judge 6);
ill-made sentences (judge 8).
- B:
very concise, well expressed (judge 7);

good sentence construction (judge 8).

- C: too wordy (judges 1, 2, 8);
well constructed and clear (judge 8).
- D: badly written, unclear (judge 2);
I didn't like the style of D very much - too chatty possibly (judge 3);
fragmented and uncohesive.
Suffers from omissions ... and verbosity (judge 8).
- E: well written, clear and understandable (judge 2);
easy to read (judge 3);
easier to read [than D], with flowing concepts (judge 6);
well constructed on the whole ... verbosity detract[s] (judge 8).

These points will be further considered in the interpretation of the second set of hypotheses.

The results for the Distinct Personality Type set are altogether different, however:

Abstract Version	A	B	C	D	E
Preference Score	8	31	22	32	27
Complexity Score (%)	38.6	70.7	58.1	72.4	60
Observed Rank	D	B	E	C	A
Expected Rank	D	B	E	C	A

In this set there is a perfect match between observed and expected ranks, signifying that the judges seem to prefer a larger amount of clause level complexity.

The ranks for the abstracts written for the first chapter of the Vickery and Vickery book are as follows:

Abstract Version	A	B	C	D
Preference Score	9	23	22	26
Complexity Score (%)	35	64.9	22.2	56
Observed Rank	D	B	C	A
Expected Rank	B	D	A	C

These results indicate that there is some measure of correspondence between the observed and expected ranks: the expected rank can be obtained by reversing the order in the first and second pairs of the observed rank. There is then slight evidence in support of the hypothesis.

The correlation is more convincing in the Chapter 2 data, where once again the hypothesis is supported:

Abstract Version	A	B	C
Preference Score	17	13	18
Complexity Score (%)	50	42.9	61.1
Observed Rank	C	A	B
Expected Rank	C	A	B

In conclusion, there are differences between the sets of abstracts in terms of the explanatory power of the hypotheses. The hypothesis was supported for the Distinct Personality Type abstracts and for the Chapter 2 abstracts. Further, there was some slight support for the hypothesis in the Chapter 1 abstracts.

Also, a weak correlation was noted between the ranks in the Tanzanian data. It would seem then that clause level complexity is a reasonably good predictor of the judges' overall preferences.

This is an interesting result because it contradicts some of the findings from the literature. In Chapter 9 above on Generic Structure, it was pointed out that Graetz claimed that abstracts were characterised by, among other things, an avoidance of subordinate clauses (1985: 125). In the case of the Information Science abstracts, on the other hand, not only is subordination very much in evidence, there is reason to suppose that, for many judges, it is a favoured strategy.

12.7 Further Hypotheses

The first hypothesis investigated the number of relations between clauses in the abstracts; this second set of hypotheses investigates the types of relation between clauses in the abstracts. This follows the implication in Halliday that grammatical intricacy comprises both the extent to which clauses combine to form clause complexes, and the means by which the clauses combine (see section 12.3 above).

12.7.1 Testing the Further Hypotheses

As stated above, clauses can be combined in one of two ways, either paratactically or hypotactically. Simultaneous with this first choice, clauses can also be combined in one of five different logico-semantic ways (=, +, x, ' or ").

If one wanted to investigate the distribution of each of these features in the data, and their possible influence on the judges' perceptions, one could postulate a set of seven hypotheses, of the form 'Abstracts with a high proportion of <feature _x> will be generally preferred over abstracts with a low proportion of <feature _x>'.

The simultaneity inherent in the language system introduces certain problems of interpretation, however; it may not be possible to assess the individual contribution of each of these variables since, for example, extended clauses can be realised either paratactically or hypotactically. Similarly, it would be dangerous to conclude that one particular judge dispreferred

hypotaxis without first investigating how hypotaxis combined with other choices in the network.

If therefore one wanted to investigate the distribution of combinations of these features in the data, and their possible influence on the judges' perceptions, one could postulate a set of ten hypotheses, of the form 'Abstracts with a high proportion of <feature _x> and <feature _y> will be generally preferred over abstracts with a low proportion of <feature _x> and <feature _y>'.

There are therefore seventeen logically distinct hypotheses which could be investigated for each of the four sets of abstracts: seven assessing the contribution of each single variable, plus the ten assessing different possible combinations of the variables. In fact, because space is limited, each of these hypotheses will not be investigated separately in deductive fashion. Rather, all the data will be analysed charting the distribution of these various features in each of the abstracts, tables of results will be presented, and it is hoped that regularities will be apparent when the observed ranks (judges' preference scores) are compared with the expected ranks (the data in the tables pertaining to clause combining). This method is more akin to analytic induction, where patterns in the data are expected to emerge (see Part Four Introduction above).

12.7.2 Problems of Analysis

A number of problems were encountered in this part of the investigation, the most difficult of which were as follows: how to differentiate between hypotaxis and parataxis, and how to draw trees for the clause complexes reliably (12.7.2.1); how to differentiate between the three different types of expansion reliably (12.7.2.2); and how to devise an appropriate scoring procedure to determine which were the most favoured means of constructing clause complexes (12.7.2.3).

12.7.2.1 Determining Interdependencies

There are two causes of difficulty concerning the notion of interdependency: the first involves the -tactic system, where very occasionally it is difficult to decide between the two options; the second has to do with univariate layering, or the drawing of trees, to represent the structure of the clause complex. These

problems will be discussed in turn.

Analysis reveals a very strong tendency towards meanings ($\alpha \wedge \beta$), rather than wordings ($1 \wedge 2$). Similarly there is a very strong tendency for projected clauses to be realised as locutions ("), rather than as ideas ('). In one or two cases, however, it is difficult to decide whether to code for hypotaxis or parataxis, since, as the example below shows, the projected clause contains elements of both meaning and wording:

Abstract Set: Distinct Personality Type
Abstract: A
Clause Complex: 13

13 ($\alpha \wedge \beta$)
13.1
Fisher states
13.2
that this fits in with the "diffident side of the common stereotype of Librarians".

As there is no provision in the framework for hybrid projections, a decision must be made between the two options. The protocol adopted was as follows: where the main verb of the projected clause is part of the meaning, code for hypotaxis (as above); where it is part of the wording, code for parataxis.

When clause complexes consist of three or more clauses, the analyst faces what is perhaps the most troublesome part of clause complex analysis: the different means of representing univariate layering. Very often there is little in the actual words themselves to signal where to attach dependencies: 'Brown and Yule and Halliday and Hasan' would be coded as $11 \wedge 12 \wedge 21 \wedge 22$, not as $1 \wedge 2 \wedge 3 \wedge 4$, since the former interpretation accords with the contextual knowledge that the four persons form two writing partnerships.

There are two related types of problem involved here: firstly, how to decide on the number of levels there should be in the dependency structure; and secondly, how to decide where to place the various attachments. The first problem is exemplified in the following clause-complex: 'It says the library is a precondition in an underdeveloped country rather than a luxury and puts forward sub-concepts, but states that options should be kept open' (see analyses below). There are *at least* three different analyses possible. The following tree implies that there are basically three elements to the clause complex, the relationship between these being one of extension, the first and last elements each containing projecting and projected clauses:

Abstract Set: Tanzania

Abstract: D

Clause Complex: 4

4 (1 α ^ 1" β ^ +2 ^ +3 α ^ +3" β)

4.1

It says

4.2

the library is a precondition in an
underdeveloped country rather than a luxury

4.3

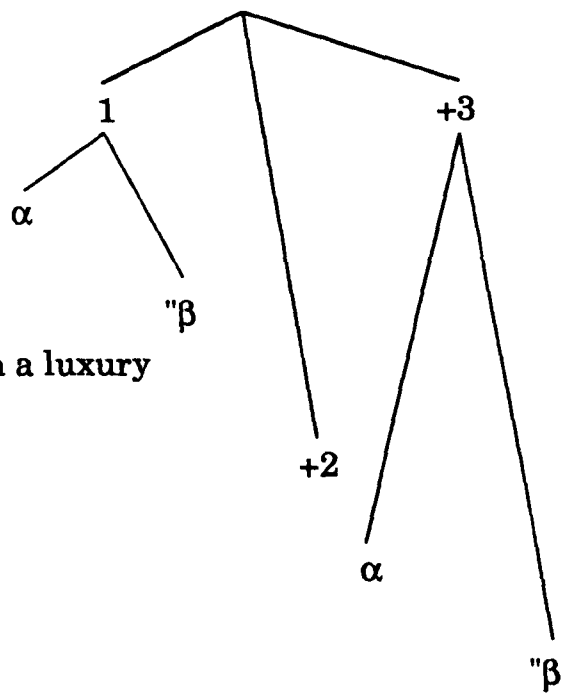
and puts forward sub-concepts,

4.4

but states

4.5

that options should be kept open.



The following tree, on the other hand, contains more bracketing, and suggests that the last two clauses are more properly viewed as a concessive corollary to the notion of the putting forward of the sub-concepts in clause 4.3:

Abstract Set: Tanzania

Abstract: D

Clause Complex: 4

4 (1 α ^ 1" β ^ +2 1 ^ +2 +2 α ^ +2 +2" β)

4.1

It says

4.2

the library is a precondition in an
underdeveloped country rather than a luxury

4.3

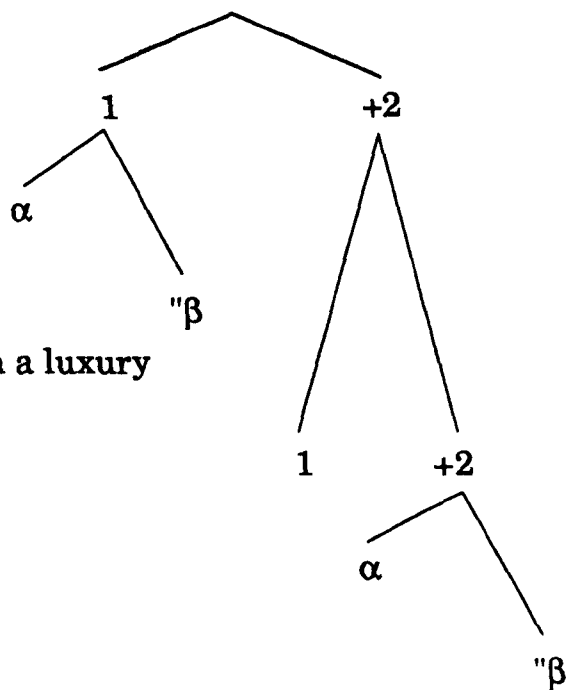
and puts forward sub-concepts,

4.4

but states

4.5

that options should be kept open.



The second problem is slightly different, in that the competing analyses do not differ in terms of their depth, but in the position of attachment of sub-components of structure. Yet another interpretation for the last example, for example, would be as follows:

Abstract Set: Tanzania

Abstract: D

Clause Complex: 4

4 (1 α ^ 1" β ^ +2 1 ^ +2 +2 α ^ +2 +2" β)

4.1

It says

4.2

the library is a precondition in an
underdeveloped country rather than a luxury

4.3

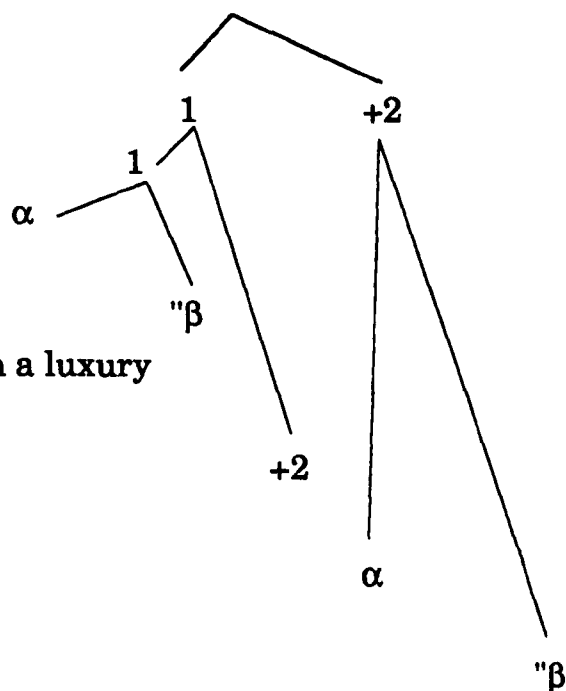
and puts forward sub-concepts,

4.4

but states

4.5

that options should be kept open.



These last two examples would receive the same depth score, but are very different in terms of dependency.

It should be emphasised that these three suggestions are not expected to exhaust all the different candidate analyses one might wish to motivate for this particular clause complex. But without knowing more of the context, it is difficult to decide between them. It is also worth noting here that the problem is really one of interpretation, and does not stem from inadequacies inherent in Halliday's descriptive framework; indeed, other types of analysis would not necessarily fare any better with this particular problem.

However, the fact that the problem is one of interpretation rather than of analysis does not lessen its seriousness given the aims and objectives of this chapter. Because of interpretive difficulties, the validity of many of the tree structures presented in Appendix 12.1 cannot not be guaranteed. For this reason it was decided, firstly, to abandon the idea of incorporating the notion of depth into scores of clause complexity, and secondly, to ensure that complexity scores were immune from this kind of problem (see 12.7.2.3 below for the derivation of a suitable scoring protocol).

12.7.2.2 Determining Logico-Semantic Relations

Within the general category of expansion, occasionally it is difficult to make a clear differentiation between elaboration and extension (= and +, respectively).

Abstract Set: Distinct Personality Type
Abstract: B
Clause Complex: 11

11 (1 ^ +2²)

11.1

The studies could be faulted on several levels,

11.2

and the pictures [produced] were so confused as [to be virtually useless] .

The example above has been coded as being an extending type of expansion, but in reality it is a moot point whether the second clause does in fact add a new element, or whether it is merely exemplifying one dimension of the studies'

flaws alluded to in the first clause. Similarly, in the following example,

Abstract Set: Vickery and Vickery, Chapter 2
Abstract: C
Clause Complex: 3

3 (1 ^ =2¹)

3.1

Organisations are interdependent,

3.2

there is a flow of information and resources between them \square keeping society going \square .

there are shades of both elaboration (the interdependence is specified as being in terms of information and resource flow) and extension (the provision of the new information that it is this relationship which keeps society going).

In most cases it was possible to ascribe categories without undue difficulty. However, it is worth pointing out that, because the logico-semantic distinctions which Halliday proposes are really quite fine, there will inevitably be some disagreements in their interpretation. Until such a time when firmer criteria are proposed, the analysis in the Appendix must therefore be regarded as tentative.

12.7.2.3 Measuring how Clauses Combine

A scoring protocol to chart each abstract's reliance on the various clause linking relations has been devised to avoid the problems of dependency discussed above. The algorithm is as follows:

for each clause of each clause complex, take the rightmost symbol (either Greek or numeric) which is prefixed by any of =, +, x, ', or ''; if there are no such prefixes, or if the clause complex consists of one single clause only, do nothing; otherwise, mark whether the prefix signifies parataxis or hypotaxis, whether projection or expansion, and if expansion, which particular sub-type.

The key to the avoidance of the dependency problem lies in the word 'rightmost' above; this ensures that only terminal symbols are scored. For example, in the following clause complex,

$$(1 \wedge =2\alpha \wedge =2=\beta\alpha \wedge =2=\beta+\beta)$$

the rightmost symbols which have prefixes are =2, =β and +β. Therefore in this particular clause complex there is one instance of a paratactic type relation, two instances of a hypotactic type relation, no instances of projection, three instances of expansion, two of which are elaborated, and one extended.

The measurement procedure is repeated such that the total numbers of the various linking relations are calculated for each abstract. To produce figures that can be fairly compared between abstracts of different length, each total is divided by the number of clause complexes in the abstract.

For the sake of an example, assume the following constitutes an entire abstract (it is actually the same example as used in section 12.4.1 above: the first six of the thirteen clause complexes of abstract version D written for the Tanzania article):

1
2
3 (1 ^ +2¹)
4 (1α ^ 1"β ^ +2 1 ^ +2 +2α ^ +2 +2"β)
5 (1 ^ +2 1α ^ +2 1"β ^ +2 +2α ^ +2 +2"β)
6 (1 ^ =2)

The rightmost prefixed symbols in this 'abstract' are:

+2¹;
"β; +2; +2; "β;
+2; "β; +2; "β;
=2.

Therefore there are:

4 examples of hypotaxis;
6 examples of parataxis;
4 examples of projection;
6 examples of expansion;
1 example of elaboration;
5 examples of extension;
4 examples of locution;
5 examples of paratactic extension;
4 examples of hypotactic locution;
and 1 example of paratactic elaboration.

Each of these scores is then normalised by dividing by the total number of clause complexes in the abstract. For example, the score for paratactic elaboration in this 'abstract' would be 0.167 (1 + 6).

Unlike the clause complexity scores discussed above, here it is not appropriate to multiply the scores by 100, and expect these to be percentage estimates. This is because it would be theoretically possible (though very unlikely) for abstracts to score more than 100% on certain variables:

- 1 (1 ^ =2)
- 2 (1 ^ =2)
- 3 (1 ^ =2)
- 4 (1 ^ =2)
- 5 (1 ^ =2 ^ =3)
- 6 (1 ^ =2)

For example, in the 'abstract' above, the rightmost prefixed symbols are:

- =2
- =2
- =2
- =2
- =2; =3
- =2.

Therefore there are 7 examples of paratactic elaboration. Multiplying by 100 would mean that the score for paratactic elaboration in this 'abstract' would be 116.7% ((7 + 6) x 100), so in the results below, figures are not given as percentages.

In the reporting of the results below (section 12.8), projections are not in fact further subdivided into locutions and ideas, since virtually all projections are realised as locutions; nor are each of the sub-types of expansion further subdivided (according to Halliday 1985: 202 - 216), since there is insufficient data available to warrant such a level of delicacy.

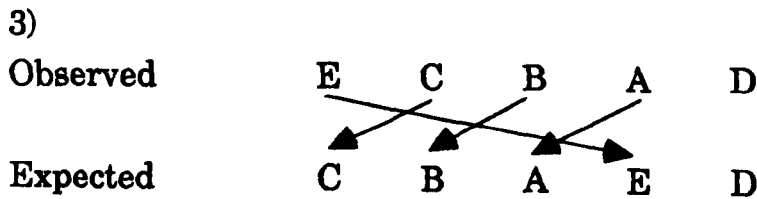
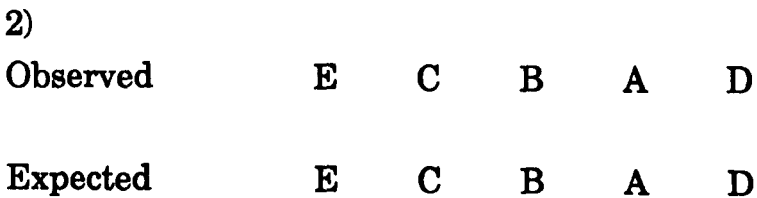
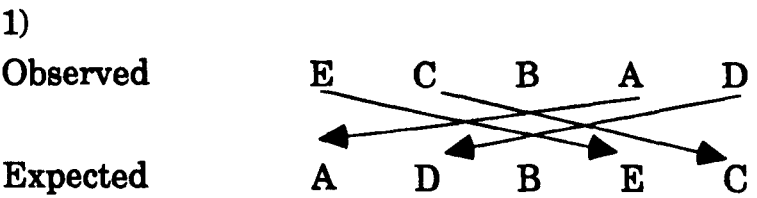
The next section presents the complete set of measurements for each set of abstracts, and attempts a reconciliation between these scores and some of the judges' evaluations, both qualitative and quantitative.

12.8 Results of Further Testing

This section is divided into four sub-sections, one for each set of abstracts. The presentation of these results proceeds inductively rather than deductively (see the Introduction to Part Four). The science of deductionism requires that hypotheses are clearly stated at the beginning of the research, and then subjected to rigorous testing; this style has been adopted for much of the work presented in Part Four. The science of analytic induction, on the other hand, expects patterns to emerge from the data; because there are so many different combinations of variables to be explored, it is more appropriate to adopt an inductive style for the remainder of this section.

The manner in which interesting features in the data come to be revealed should be as principled as possible, however; induction is not an excuse for 'only seeing what you want to see'. The following strategy has been developed for calling attention to the most interesting patterns in the data.

The 'shifts away score', first discussed in Chapter 10 (see section 10.6 above), provides a convenient means of comparing the goodness of fit between different pairings of observed and expected ranks. For example, consider the following three sets of observed and expected ranks:



When there are two sets of five items, 12 is the highest possible shifts away score, as exemplified by pairing number 1. When there is a perfect match

between ranks, the shifts away score is 0, as exemplified by pairing number 2. Pairing number 3 represents a mid-range match, scoring 6 shifts away.

For the purposes of this chapter, the most interesting pairings will be taken to be those which are either obviously positively correlated, or obviously negatively correlated. Thus, 'interestingness' can be quantified in the following way:

12	11	10	9	8	7	6	5	4	3	2	1	0
-ve correlation				uninteresting				+ve correlation				

This will be used to filter out the most interesting pairings between the ranks expected by the various clause combining variables and the judges' observed ranks concerning the five abstracts summarising the Tanzania article (see section 12.8.1 below) and the five summarising the Distinct Personality Type article (see section 12.8.2 below).

Because there were only four abstracts written for the first Vickery and Vickery book chapter and eight is therefore the maximum shifts away score, the following filter will be used in section 12.8.3:

8	7	6	5	4	3	2	1	0
-ve correlation			uninteresting			+ve correlation		

Because there were only three abstracts written for the second Vickery and Vickery book chapter and four is therefore the maximum shifts away score, the following filter will be used in section 12.8.4:

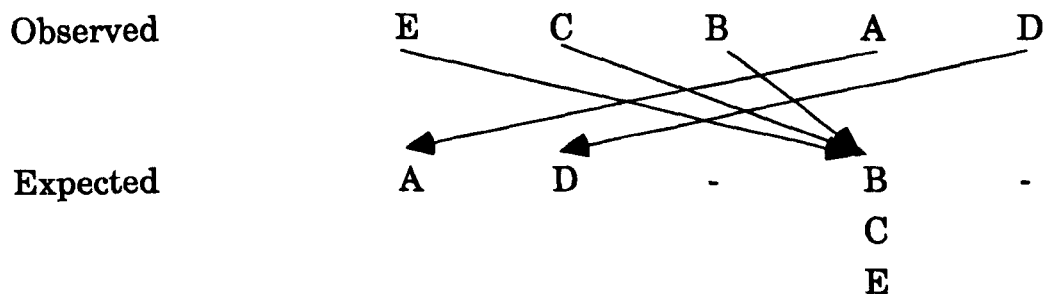
4	3	2	1	0
-ve correlation	uninteresting			+ve correlation

12.8.1 Tanzania Results

The scores for the Tanzanian set of abstracts are shown overleaf:

Observed Rank	E	C	B	A	D	Shifts Away Score
	Tanzania Version A	Tanzania Version B	Tanzania Version C	Tanzania Version D	Tanzania Version E	
Paratactic Elab	0.053	0	0	0.032	0	12
Paratactic Ext	0.158	0.286	0	0.29	0.045	12
Paratactic Enh	0	0	0	0	0	6
Paratactic Proj	0	0	0	0	0	6
Hypotactic Elab	0.105	0	0	0	0	9
Hypotactic Ext	0	0	0.091	0.032	0.045	6
Hypotactic Enh	0.053	0	0.091	0.064	0.091	5
Hypotactic Proj	0.053	0.143	0.091	0.161	0.227	6
Total Parataxis	0.21	0.286	0	0.323	0.045	12
Total Hypotaxis	0.21	0.143	0.273	0.258	0.364	4
Total Elab	0.158	0	0	0.032	0	12
Total Ext	0.158	0.286	0.091	0.323	0.091	12
Total Enh	0.053	0	0.091	0.064	0.091	5
Total Projection	0.053	0.143	0.091	0.161	0.227	6

Shifts away scores are given for each variable in the right hand column. For example, the shifts away score for the Paratactic Elaboration variable would be calculated as follows:



A has the highest score (0.053), so this is placed first. D has the second highest score (0.032), so this is placed second. There is then a tie, so the three abstracts are placed in joint fourth position. The total score is given by adding together the five shifts which must be made to bring the observed and expected ranks into alignment: 3 shifts for E; 2 shifts for C; 1 shift for B; 3 shifts for A; and 3 shifts for D, making a total of 12.

The most interesting of the shifts away scores are shown in bold type in the rightmost column. These are:

- Paratactic Elaboration (negatively correlated);
- Paratactic Extension (negatively correlated);
- Hypotactic Elaboration (negatively correlated);
- Total Parataxis (negatively correlated);
- Total Hypotaxis (positively correlated);
- Total Elaboration (negatively correlated);
- Total Extension (negatively correlated).

It seems that the judges are marking down abstracts which rely heavily on paratactic means of linking clauses, paratactic extension and paratactic elaboration in particular.

On the other hand, it seems that the judges prefer those abstracts which rely heavily on hypotaxis, particularly when it is non-elaborated.

As pointed out in section 12.6 above, the Tanzanian set attracted many qualitative comments which it was thought might be explained by the type of analysis investigated in this chapter. Some of these appear below.

Concerning version A of the Tanzanian set:

judge 2 said 'quality of writing not too good';

judge 3 said 'too disjointed. I prefer proper sentences';

judge 5 said 'very jerky portrayal of concepts. Didn't flow at all'.

Perhaps some of these observations are linked to the fact that version A had the lowest overall clause level complexity score (59.3%).

Concerning version B of the Tanzanian set:

judge 1 said 'totally point-like to read';
judge 4 said 'excerpts not an abstract'.

Interestingly, B scores relatively highly for paratactic extension, a dispreferred strategy. One suggestion which might explain these two qualitative remarks might be that the abstractor has taken many of the original words and phrases from the source material, concatenated them using simple 'and'-type links, and has not paid sufficient attention to the coherence of the abstract as text in its own right.

Concerning version C of the Tanzanian set:

judge 1 said 'too wordy';
judge 2 said 'more wordy but a fair summary';
judge 8 said 'wordy in places'.

Version C scored second lowest on clause level complexity (60%). Without wishing to claim too much, it is perhaps worth pointing out that, following Halliday's observation that writing tends to be grammatically simple whereas speech tends to be grammatically intricate, version C might be more written than spoken, as it were; and this, together with the fact that C is the only one of the set not to contain any examples of parataxis, might partially explain its perceived 'wordiness'.

Concerning version D of the Tanzanian set:

judge 2 said 'badly written, unclear';
judge 3 said 'I didn't like the style of D very much - too chatty possibly'.

The 'chattiness' referred to here might be explained by the fact that D contains the highest proportion of extended clauses (0.323), most of which are realised paratactically. This is in contrast to the 'wordy' C, which has only 0.091

extensions, all of which are realised hypotactically. The suggestion here is that paratactic extension, in particular, is a combination more frequently associated with speech than writing.

To recap: when judges complain of an abstract's 'wordiness', the suggestion is that that abstract will be less grammatically intricate (i.e. will have fewer clause combining relations); when they complain of an abstract's 'chattiness', on the other hand, the suggestion is that that abstract will be more grammatically intricate, especially in terms of paratactic extension. To please all the judges, then, it is as if abstracts must combine some of the features of spoken language with some of the features of written language.

Concerning version E of the Tanzanian set:

judge 2 said 'well written, clear and understandable';
judge 3 said 'E is informative, easy to read and has a beginning, middle and end'.

E contains relatively little parataxis, and a high proportion of enhancement. Although it may well have been only a subconscious association, the fact that the judge has used the lexicalisation 'informative' is very interesting. Once again, without wishing to over-claim, it is tentatively suggested that informative abstracts might make greater use of enhancement than would indicative abstracts, since in an informative abstract there is greater scope for 'embellishing' (Halliday's way of describing the function of enhancement, 1985: 197) the information provided.

In short, then, the judges seem to prefer those versions which contain hypotaxis rather than parataxis, and enhancement rather than elaboration or extension. Projected material does not seem to affect the judges one way or the other.

12.8.2 Distinct Personality Type Results

The scores for the Distinct Personality Type set of abstracts are shown overleaf:

Observed Rank	D	B	E	C	A	Shifts Away Score
	DP Type Version A	DP Type Version B	DP Type Version C	DP Type Version D	DP Type Version E	
Paratactic Elab	0.085	0.074	0.167	0	0	12
Paratactic Ext	0.068	0.185	0	0.21	0	4
Paratactic Enh	0	0	0	0	0	6
Paratactic Proj	0	0	0	0.053	0	4
Hypotactic Elab	0	0	0	0	0	6
Hypotactic Ext	0	0	0.056	0	0.111	10
Hypotactic Enh	0.034	0.111	0.056	0.158	0.222	4
Hypotactic Proj	0.085	0.185	0	0.158	0	6
Total Parataxis	0.152	0.259	0.167	0.263	0	4
Total Hypotaxis	0.119	0.296	0.111	0.316	0.333	6
Total Elab	0.085	0.074	0.167	0	0	12
Total Ext	0.068	0.185	0.056	0.21	0.111	2
Total Enh	0.034	0.111	0.056	0.158	0.222	4
Total Projection	0.085	0.185	0	0.21	0	4

The most interesting of the shifts away scores (emboldened in the rightmost column) are:

- Paratactic Elaboration (negatively correlated);
- Paratactic Extension (positively correlated);
- Paratactic Projection (positively correlated);
- Hypotactic Extension (negatively correlated);
- Hypotactic Enhancement (positively correlated);
- Total Parataxis (positively correlated);
- Total Elaboration (negatively correlated);
- Total Extension (positively correlated);
- Total Enhancement (positively correlated);
- Total Projection (positively correlated).

It seems that the judges prefer those abstracts which rely heavily on paratactic means of linking clauses, particularly when extended or projected. This is in stark contrast to the judges' preferences concerning the Tanzanian set, where parataxis and extension seemed to be dispreferred clause linking strategies.

Also it seems that the judges prefer those abstracts which rely heavily on projection, particularly when it is realised paratactically. Again, this differs from the Tanzanian results, where projection did not seem to make much difference to the judges' opinions.

The fact that there is such a difference between these first two sets is interesting, and will be returned to in section 12.9 below.

There is altogether less material in the qualitative data which brings to mind issues of readability, but it is worth mentioning that Judge 3 said that version B was 'fairly easy to read', while Judge 7 said that version D 'reads well, but seems to be short on detail'. Overall these were the two most preferred abstracts, and contain by far the greatest parataxis and projection. In the abstracts projection serves to ascribe wordings and meanings to particular authors, and perhaps the readers appreciate being told who is saying what, as it were. The disadvantage of using a high proportion of projection is that there is concomitantly less space to impart the raw knowledge, a fact which might explain why D, the abstract with most projected material, was thought 'short on detail'.

12.8.3 Vickery and Vickery, Chapter 1 Results

The scores for the first book chapter set are shown overleaf:

Observed Rank	D	B	C	A	Shifts Away Score
	V&V1 Version A	V&V1 Version B	V&V1 Version C	V&V1 Version D	
Paratactic Elab	0	0	0	0.053	2
Paratactic Ext	0	0.111	0.067	0	4
Paratactic Enh	0	0.111	0	0	4
Paratactic Proj	0	0	0	0	4
Hypotactic Elab	0	0	0	0.053	2
Hypotactic Ext	0	0.111	0	0.053	3
Hypotactic Enh	0.133	0.185	0	0.158	4
Hypotactic Proj	0	0.074	0	0.105	1
Total Parataxis	0	0.222	0.067	0.053	4
Total Hypotaxis	0.133	0.296	0	0.368	2
Total Elab	0	0	0	0.105	2
Total Ext	0	0.148	0.067	0.053	4
Total Enh	0.133	0.296	0	0.158	4
Total Projection	0	0.074	0	0.105	1

The most interesting of the shifts away scores (emboldened in the rightmost column) are:

- Paratactic Elaboration (positively correlated);
- Hypotactic Elaboration (positively correlated);
- Hypotactic Projection (positively correlated);
- Total Hypotaxis (positively correlated);
- Total Elaboration (positively correlated);
- Total Projection (positively correlated).

It seems that the judges prefer those abstracts which rely heavily on hypotactic means of linking clauses, on projections, and on elaborations. This is similar to the judges' preferences concerning the Distinct Personality Type set, where projection was also the preferred logico-semantic linking strategy. However it is different from the Distinct Personality Type set in that hypotaxis has replaced parataxis as the preferred -tactic alternative.

Again, there is comparatively little in the qualitative data to be commented upon, although the following point can be noted. Version B, the abstract judged 'easy to read' (judge 3) and thought to read well (judge 7) has the highest proportion of parataxis by far of the four versions, and the second highest value for projection. Parataxis and projection were also found to correlate with ease of reading in the Distinct Personality Type set.

12.8.4 Vickery and Vickery, Chapter 2 Results

The scores for the second book chapter set are shown overleaf:

Observed Rank	C	A	B	Shifts Away Score
	V&V2 Version A	V&V2 Version B	V&V2 Version C	
Paratactic Elab	0	0.042	0.067	2
Paratactic Ext	0.062	0.125	0.067	4
Paratactic Enh	0	0.042	0.133	2
Paratactic Proj	0	0	0	2
Hypotactic Elab	0.062	0	0	3
Hypotactic Ext	0	0	0	2
Hypotactic Enh	0.062	0.125	0.267	2
Hypotactic Proj	0	0	0	2
Total Parataxis	0.062	0.208	0.267	2
Total Hypotaxis	0.125	0.125	0.267	1
Total Elab	0.062	0.042	0.067	0
Total Ext	0.062	0.125	0.067	4
Total Enh	0.062	0.167	0.4	2
Total Projection	0	0	0	2

The most interesting of the shifts away scores (emboldened in the rightmost column) are:

- Paratactic Extension (negatively correlated);
- Total Elaboration (positively correlated);
- Total Extension (negatively correlated).

This set of results is altogether harder to interpret, partly because there were only three versions written. The only one of the scores to predict the judges' observed preferences perfectly was that for elaboration. This set is unusual in that it is the only one not to contain any projected material.

In short, no clear conclusions can be drawn from these quantitative results, although the overall extent of clause level complexity was found to perfectly predict judges' preferences (see section 12.6 above).

The qualitative data reveals that judges 3 and 4 thought that version B was 'disjointed', an observation which might partially be accounted for by its low score for clause level complexity. Judge 3 thought that version C was 'very clear', and judge 5 considered it 'by far the most readable'; the fact that C contained the highest proportion of parataxis (0.267) squares with one of the findings concerning reading ease discussed in 12.8.2 above.

Judge 1, on the other hand, thought that the same abstract was, on the contrary, 'very difficult to read'. This last observation prompts two last admonishments: firstly, it must be remembered that the quantitative data discussed in this chapter pertains to the judges' *overall* preferences, preferences which themselves contain a good deal of variation; and secondly, this may in turn lead to the hypothesis that each of the impressionistic terms used by the informants (such as 'wordiness', 'disjointedness', etc.) are better described by a particular configuration of a cluster of linguistic features, than by any single variable.

12.9 Further Discussion of the Results

In an interesting article published in 'Word', Fries (1986: 13) says:

'There has been a long tradition of searching in texts for formal characteristics which will correlate with ease of reading. Such attempts include the linguistically naive readability formulae which measure syntactic complexity by finding the mean lengths of sentences or T-units, and measure vocabulary difficulty by determining the number of infrequent or long words (Klare: 1963). They also include such linguistically sophisticated notions as the depth hypothesis proposed by Victor Yngve some twenty years ago (Yngve 1960). I will claim that, even though meanings are realized in linguistic form, mere counting of the language forms contained in a text will not lead to useful judgments of the readability or coherence of that text'

It is submitted that there is evidence in this chapter to suppose that counting language forms can indeed be useful, provided that such internal measures are used in conjunction with external ones (see Chapter 1 above for the

distinction between the two). One of the most basic arguments promulgated by this thesis is that external measures, readers' success judgements in other words, should always be seen to be primary; internal measures are taken in order to explicate external measures.

In this chapter it was discovered that overall clause level complexity predicted reader preference relatively well for all abstracts. Further, in the investigation of the particular types of linking relation used, a sharp divergence was noted between the Distinct Personality Type set and the Tanzanian set: the best Distinct Personality Type abstracts preferred parataxis and extension, whereas the best Tanzanian abstracts dispreferred them. Also, the best Distinct Personality Type abstracts relied heavily on projection, as did those for the first book chapter, whereas projection did not seem to predict informants' preferences for the Tanzanian set.

Without wishing to over-claim, it is suggested that judges may have ranked the two sets of abstracts - Distinct Personality Type and Tanzania - according to different criteria: it seems plausible that judges were expecting something more akin to an informative abstract for the Tanzanian source text, but something more akin to an indicative abstract for the Distinct Personality Type source text (see Chapter 5 above for the distinction).

An indicative strategy is arguably more appropriate for the Distinct Personality Type source text because it is a review article, citing the work and opinions of many different researchers. A high proportion of projection might therefore be expected because projecting material is better suited to situating the work in a broader interpersonal framework, giving fuller details of the people involved in the work, and its relation to the readership.

An informative strategy is arguably more appropriate for the Tanzania source text, because it is more like a problem - solution type of text. A high proportion of enhancement might therefore be expected because enhancement is better suited to providing certain types of ideational meaning, particularly involving the logic or circumstances underpinning the argument presented in the work.

Abstracts written for the two book chapters lie somewhere between the informative and the indicative: however, if greater projection is indeed a characteristic of the indicative abstract, it is hypothesised that readers would have preferred an indicative abstract for the first chapter of the book. This would tally with Borko and Bernier's observation (1975: 16) that 'indicative abstracts are frequently used for describing discursive or lengthy texts, such as broad overviews'.

12.10 Conclusions

The following conclusions are to be drawn from this chapter:

- Grammatical intricacy can be measured in two ways: the first calculates the extent to which clauses combine in a text to form larger structural units; the second calculates a text's reliance on each of the different ways in which clauses can be joined together.
- Both these measures have been shown to be useful in explaining informants' intuitions.
- Overall, judges prefer abstracts with higher clause complexity scores.
- However, some types of complexity are preferred over other types of complexity.
- The following table summarises the relationships between the informants' preferences and the different ways in which clauses can be linked together (crosses denote negative correlations and ticks denote positive correlations):

	Tanzania	Distinct Personality Type	V & V, Chapter 1	V & V, Chapter 2
Paratactic Elab	✕	✕	✓	.
Paratactic Ext	✕	✓	.	✕
Paratactic Enh
Paratactic Proj	.	✓	.	.
Hypotactic Elab	✕	.	✓	.
Hypotactic Ext	.	✕	.	.
Hypotactic Enh	.	✓	.	.
Hypotactic Proj	.	.	✓	.
Total Parataxis	✕	✓	.	.
Total Hypotaxis	✓	.	✓	.
Total Elab	✕	✕	✓	✓
Total Ext	✕	✓	.	✕
Total Enh	.	✓	.	.
Total Projection	.	✓	✓	.

For example, there is a positive correlation between the total amount of projection and the judges' ranks for both the Distinct Personality Type and the V&V, Chapter 1 abstracts. This would suggest that the judges respond well to abstracts which feature a relatively large amount of projected material.

- Correlations vary across the different source texts.
- It is suggested that the distribution of clause relations may provide a useful way of differentiating between sub-types of abstract. One working hypothesis might be that indicative abstracts feature relatively large amounts of projection, while informative abstracts feature relatively small amounts of parataxis.

13

Analysis of Themes

13.1 Introduction

This chapter assesses the claim that a text's thematic organisation may predict *to a certain extent* how successful that text will be perceived to be. Having motivated the choice of Theme as one means of explaining some of the informants' opinions (section 13.2), a number of definitions of theme are provided, and their theoretical background given (section 13.3). A selection of initial hypotheses relating theme to success are then formally stated in section 13.4. These are taken mainly from the literature and provide a useful starting point for the analysis. Section 13.5 is a long section in which the testing of the initial hypotheses is described; the section incorporates discussion of problems encountered during the analysis phase.

Problems arising both from the peculiarities of the data and from the imperfections of the descriptive framework had to be faced, and so a section is devoted to each. In the discussion of the decisions which were taken to make the analysis appropriate to the abstracting context, particular attention is paid to those areas where it was considered necessary to depart from orthodox thematic analysis. In fact, the type of analysis espoused by Halliday 1985 is largely followed with some modifications: these chiefly concern the delimitation of theme, and its treatment in non-main clauses. Since the discussion of the problems stretches to a number of pages, a short summary is given outlining the main decisions that were taken prior to analysis. An explanation of the various codes used to label the texts is also provided.

A number of quantitative measures have been derived from the theme analysis, which facilitate the testing of explicit hypotheses, hypotheses being taken mainly from the literature. Results for each hypothesis are detailed, together with the relevant tables showing the numerical findings. These tables have been derived from the analyses presented in the Appendices: Appendix 13.1 provides a full thematic interpretation of all seventeen of the Information

Science abstracts; and Appendix 13.2 tabulates the themes grouped according to type, whether topical, interpersonal, or textual, categories taken from Halliday's third chapter in IFG.

Sections 13.6 and 13.7 investigate three further hypotheses. These owe less to previous research reported in the literature, and stem from consideration of patterns evident in the abstracts data. This last set of hypotheses concerns the distribution of different types of topical theme. A new kind of topical theme is derived - discoursal theme (see section 13.6) - and it is hypothesised that the tripartite division of topical themes into informational, interactional and discoursal can help to explain some of the judges' preferences. Appendix 13.3 tabulates the three types of topical theme in each of the Information Science abstracts. Once again, quantitative results are used to examine the validity of the various hypotheses.

In section 13.9, remarks are made assessing how well the different types of topical theme explain judges' preferences, and the correspondence between these and the type of the abstract is noted. Finally, brief conclusions are given in section 13.10.

13.2 Why Analyse Themes?

One of the most interesting findings reported in Chapter 8 above was that judges often seem to be less concerned about what was said in the abstracts, and more concerned about how it was said, issues of readability being particularly to the fore.

With this in mind, it was considered potentially instructive to investigate how choice of theme affects readers' perceptions of quality. There were two main reasons for this.

First and foremost, there is much in the informants' qualitative comments which motivate thematic analysis. The following two verbatims are particularly pertinent in this regard:

'better ordered but still not clear as to salient points of article'
(judge 2 on Distinct Personality Type, version C);

'difficult to scan and absorb'
(judge 8 on Distinct Personality Type, version C);

These two qualitative remarks particularly motivate the choice of thematic analysis, since the judges are complaining about the absence of precisely those qualities Martin claims are engendered by certain configurations of thematic patterning (see 13.3.2 below). One of his observations (1986a: 39 - 40, cited in Berry 1989) is that the skilful arrangement of topical items in first position can aid skim reading and promotes easy navigation through the text.

Secondly, in her own study of abstracts, a study which shares many of the underlying aims and objectives of this research, Drury 1989b found that particular configurations of thematic patterning differentiated well between abstracts written by first and second language students (1989b: 163 - 175, see also Chapter 3 above).

The next section (13.3) provides some theoretical background to the theme concept, prior to the testing of some hypotheses taken mainly from the literature (sections 13.4 and 13.5).

13.3 Theoretical Background

This section will provide a brief introduction to Theme: Halliday's theory is discussed in section 13.3.1; recent developments of his work are then reviewed in section 13.3.2.

13.3.1 Halliday's Approach

Other schools of linguistics have notions which are roughly analogous to theme, but followers of Systemic linguistics have, in particular, spent much time and effort discussing (and arguing about) the concept. In fact, the theme/rheme division originated in Prague School linguistics, and is at the heart of 'thematic progression', an idea geared to the interpretation of whole texts (see for example Firbas 1964, Danes 1974). In contrast, Halliday's treatment of theme is primarily concerned with individual clauses; it has been left to other researchers to consider how such a mechanism has relevance for the level of text (see for example Fries' 'method of development', 1983).

Theme is described most comprehensively in the second part of Halliday's 'Notes on Transitivity and Theme in English' (1967). For Halliday, the theme is 'what comes first in the clause ... what is being talked about, the point of

departure for the clause as a message' (1967: 212), and 'is that with which the clause is concerned' (1985: 38). In English, the theme is represented by first position in the clause, and signals to the hearer the speaker's conception (1967: 212) of "what I am talking about" (or "what I am talking about now").

Halliday's chapter on theme in IFG (1985: chapter 3) gives many more examples and provides a useful schematic means of analysing theme, but is basically a restatement of the earlier account. New to IFG are the following features: what were referred to as 'composite' themes, and treated as being 'problematic' (1967: 218 - 223), reappear in the later version, and are handled more exhaustively as 'multiple themes' (which in turn enable Halliday's most precise definition yet: 'The Theme of any clause, therefore, extends up to (and includes) the topical Theme', 1985: 56); advice is given on how to handle dependent, minor and elliptical clauses (1985: 61 - 64, though not without problems, see below); some examples are coded twice, once at the level of the clause, and once at the level of the clause complex; and some interesting associations are made between theme choice and grammatical metaphor. The tripartite division of multiple theme into textual, interpersonal (two optional thematic elements) and topical (the obligatory thematic element, at least in non-elliptical, main clauses) is a particularly interesting and powerful development, and is used extensively in what follows. This is not to say of course that there are no problems with theme. There are many, some of a theoretical nature, and some practical; problems confronting the person wanting to use theme as an analytical tool are reviewed below in sections 13.5.1 and 13.5.2.

The two most basic theoretical problems are knowing what exactly theme is, and what exactly it does. Though definitions abound in the literature (see Fries and Francis 1990 for various formulations), both Hudson (1986: 798) and Huddleston (1988: 158 - 162) complain that it is at best only vaguely defined. So, while it may be possible to provide a formal realisation (although not without some problems, as will be seen below), it is altogether more difficult, they would argue, to see how some elements of the clause, which are identified as being thematic, act as a marker for what the clause is about; Hudson, for example, questions how subordinating 'that' conjunctions can of themselves be considered to signal 'aboutness'.

Even discounting such dependent clauses, and interrogative and imperative clauses likewise, all of which are notoriously troublesome, Huddleston's arguments (1988: 158) concerning the alleged topicality of the initial constituent in existential 'there' clauses, and in examples such as 'Nothing

will satisfy you' and 'You could buy a bar of chocolate like this for 6d before the War', are convincing, and necessitate some refinement of the framework.

Ideally, this thesis would attempt to solve all such theoretical problems, and then proceed to consider the more practical problems encountered when analysing naturally occurring texts. However, it is submitted that solving both kinds of problem is beyond the scope of this thesis, principally because theme is only one type of analysis investigated. Because this research is intended to be applied rather than theoretical, a more pragmatic approach will be assumed, and, while admitting that the theory underlying theme is not without its difficulties, more practical issues will be concentrated upon in the attempt to actualise the insights thematic analysis promises to provide.

13.3.2 Developments of Halliday's Approach

In spite of the many allied problems, a number of linguists have found theme to be both a usable and useful construct. There have been two particularly exciting applications of thematic analysis. One attempts to show how theme can be used as one means of differentiating between writing belonging to different genres (see Stainton forthcoming (c), for example, or Francis 1989), the other how it can be used to differentiate between texts judged to be of differing levels of success (one early example being Fries and Fair 1978, who tried to show how good writers favour particular patterns of cohesive relations between items in their themes).

It is this second type of application which is particularly relevant for the goals of this research. Martin, in particular, has for the last few years been interested in how and why some texts are perceived to be more successful than others (1985, 1986a), investigating to what extent theme, among many other types of analysis he employs, is correlated with success (see for example his comments on the different thematic means of relating sections of text to their context, 1986a: 38 - 40; and his claim that different types of topical theme help to signal different ideological authorial stance in texts, 1986b: 240 - 241).

To date, the bearing that theme may have on the perceived success of written discourse has been most extensively investigated by Berry (1989, 1990 and forthcoming). Following Halliday's third chapter of IFG quite closely, she shows (1990) how a consideration of themes classified into textual, interpersonal and topical groups can account to some degree for various informants' judgements concerning the relative success of two versions of a

letter from a mail order company written by two different people, but saying similar things. In her 1989 article, which came after the 1990 presentation logically but not chronologically, she demonstrates a correspondence between the distribution of certain types of topical theme in the writing of school children and her own intuitions concerning the degree to which these various texts constitute successful instances of the target genres.

Berry was careful to preface her 1990 presentation with a disclaimer: theme alone should not be expected to say everything there is to be said about success in writing; she did however suggest that it was certainly one avenue worth exploring. The point is well taken in this study. It would be most unlikely if theme were to be seen to account for all the variation in the judges' rankings; just as it would be unrealistic to suppose that the judges' qualitative remarks reflect an exclusive preoccupation with thematic issues. The argument in this chapter is merely that theme is a relatively good predictor of the informants' success judgements.

This claim will now be addressed.

13.4 Statement of Initial Hypotheses

In this section a number of hypotheses will be presented. These have been taken from the literature, and will be investigated in the next section.

Hypotheses concerning textual themes will be considered first.

Comparatively little has been said about textual theme; in common with many thematic interpretations, Halliday's own commentary on the Stevenson text (1985: 67) reflects primary concern with topicality. Drury, on the other hand, has investigated textual themes more systematically in her study of abstracting (1989b: 163 - 175), and concludes that the poorer summaries in her collected data exhibit noticeably higher proportions of textual themes. This then shall be the first hypothesis:

• Hypothesis 1:

Abstracts with fewer textual themes will be generally preferred over abstracts with more textual themes.

Drury charts only one instance of interpersonal theme in her three summaries: the native speaker's version, which was judged the best overall, contained a mention of 'above all', which Drury treats as being interpersonal.

Interestingly, one of Stainton's hypotheses (forthcoming (c)) goes against this admittedly rather small evidence. She hypothesised that successful instances of articles would be characterised by generally fewer interpersonal themes than in unsuccessful instances. She reports that this hypothesis was only partially supported, and concluded that a low frequency of interpersonal themes appeared to be correlated with one particular sub-type of article, the scientific article, arguing that, in general, scientific articles would be expected to be written in an objective style, one in which the author's voice was relatively silent. Given that objectivity is one of the first skills novice abstract writers are taught, it can be hypothesised that:

- **Hypothesis 2:**

Abstracts with fewer interpersonal themes will be generally preferred over abstracts with more interpersonal themes.

Turning now to topical themes, little of interest might be expected to emerge from a consideration of the distribution of topical elements in main clauses, since, according to Halliday's definition, every main clause contains one topical theme. In the data, however, main clauses do not always necessarily contain a topical theme; some are elided. Similarly non-main clauses do not always contain topical themes. If Martin is correct in his suggestion that topical themes can aid navigation through texts, then it might be expected that abstracts with fewer explicit topical themes entail greater cognitive load in the mind of the reader, and may even cause frustration. It is therefore hypothesised that:

- **Hypothesis 3:**

Abstracts with more topical themes will be generally preferred over abstracts with fewer topical themes.

In addition, both Francis 1989 and Drury consider the lexical density of topical elements; in line with Drury's observation that the summary which received the best overall marking had the most lexically dense topical themes, it can be hypothesised that:

- **Hypothesis 4:**

Abstracts which have longer topical themes will be generally preferred over

abstracts which have shorter topical themes¹.

As well as these hypotheses which relate to the type of theme, whether textual, interpersonal, or topical, it can be investigated whether the option of marked versus unmarked theme (Halliday 1985: 45) has any bearing on judges' rankings. Stainton (forthcoming (c)) has examined the distribution of marked theme in articles, and reports that articles judged more successful make more use of marked themes than articles judged less successful. It can therefore be hypothesised that:

- **Hypothesis 5:**

Abstracts which have marked themes will be generally preferred over abstracts which have fewer or no marked themes.

Similarly she found a higher number of clausal themes seemed also to be correlated with the more successful articles. It can therefore be hypothesised that:

- **Hypothesis 6:**

Abstracts which have clausal themes will be generally preferred over abstracts which have fewer or no clausal themes.

These few hypotheses will serve as a useful starting point in the investigation of theme as a potential determinant of success. The next section discusses the decisions that had to be taken in order to overcome problems in devising a valid and reliable means of testing these first hypotheses.

13.5 Testing the Initial Hypotheses

This section is rather long and involved, since it describes a number of the problems which plagued the investigation. Section 13.5.1 below discusses features of the data which caused difficulty in the testing of the hypotheses,

¹Although longer themes are *not necessarily* more lexically dense than shorter themes, work carried out for a previous version of this chapter established that, in practice, the two different measures - average number of words per theme and average LD per theme (see Chapter 10 above for a discussion of LD) - yielded the same expected ranks for the abstracts data. For the sake of simplicity, therefore, only the former measure is reported in this new version.

and immediately following, section 13.5.2 discusses features of the descriptive framework which likewise made the task of analysis somewhat less than straightforward. Because the discussion of these problems stretches to a number of pages, a summary of the decisions taken in the light of these problems is given in section 13.5.3, together with a key to the symbols used in the thematic analysis. Full versions of the abstracts, each thematically annotated following Halliday's box notation in his theme chapter (1985: chapter 3), are given in Appendix 13.1. Discussion of how the necessary quantitative measures were derived is then provided towards the end of this section, and results and conclusions given.

13.5.1 Problems of Analysis Arising from the Nature of the Data

This section describes some difficulties stemming from peculiarities in the data; difficulties stemming from certain inadequacies in the theory are dealt with in the next section.

Because of the way in which certain of the abstracts were written, some of the data is resistant to straightforward grammatical interpretation. One of the generic peculiarities of the abstract is that space is generally at a premium, and it is perhaps this constraint which has prompted some of the abstract writers to adopt something akin to a note-taking style. Particularly troublesome with regard to theme are those examples which have no main verb (the importance of the main verb will be discussed below in section 13.5.2.1). A selection of these appear below:

Abstract Set: Distinct Personality Type

Abstract: C

Clause Complex: 6

6

Librarians image ambiguous as both severe and diffident.

Abstract Set: Tanzania

Abstract: A

Clause Complex: 4

4

Introduction of theories borrowed from abroad,
and fitted into social context.

Abstract Set: Vickery and Vickery, Chapter 1

Abstract: A

Clause Complex: 1

1

1) Information Science: Emergence and Scope

The study of communication of information in society.

These examples present the analyst with the following dilemma: whether to 'guess' the intended main verb, and code for theme accordingly; or whether to leave such data unanalysed.

The first of the three examples above might perhaps have been analysed as having 'Librarians image' as its theme, the rest being regarded as a rhematic attribute to a missing relational verb. In the second and third examples, however, it seems more likely that the author deliberately chose not to include a main verb, as in the following example:

Abstract Set: Tanzania

Abstract: C

Clause Complex: 1

1

An evaluation of the performance of the public library system in Tanzania, attempting to identify the problems and limitations imposed by the adverse economic conditions and the defects of organisation and direction within the service.

In the interests of consistency throughout the analysis, the decision was taken not to provide a thematic interpretation for any clause complex missing a main verb.

There is evidence in the qualitative judgements to suggest that informants generally prefer whole sentences. This is especially true of version A of the 'Distinct Personality Type' set of abstracts, which is universally dispreferred by the judges. This abstract contains much that is grammatically ill-formed, making it extremely difficult to analyse thematically. Therefore, although for the sake of completeness an attempt has been made to chart its themes, it must be remembered that the analytical framework was not designed to take account of this kind of nonstandard input, and so the reliability of the thematic interpretation cannot be guaranteed.

13.5.2 Problems of Analysis Arising from the Imperfections of the Descriptive Framework

This section describes those decisions which were taken in order to minimise certain difficulties arising from a number of theoretical problems inherent in the descriptive apparatus. The first of these (section 13.5.2.1) is quite general, and challenges Halliday's interpretation of the boundary between theme and rheme. The second (section 13.5.2.2) stems from the fact that in the abstracts data there is a lot of information which is realised in non-main clauses. Although such clauses are often left unanalysed by researchers investigating other types of writing, it is argued that the abstracts data must be analysed more fully, otherwise significant contributions to thematic organisation would go unnoticed. Section 13.5.2.3 discusses problems of ellipsis, 13.5.2.4 problems of grammatical metaphor, and finally, some remarks concerning marked Theme are made in section 13.5.2.5.

13.5.2.1 The Delimitation of Theme

In this study a number of departures have been made from the way theme is analysed in Halliday 1985. The first of these concerns delimitation, or where to draw the boundary between theme and rheme.

Enkvist's paper on 'theme dynamics' begins with the opening sentence (1973: 127):

In a bird's eye view, my paper will try to relate two streams of modern linguistics to the study of literary texts.

Under the Hallidayan view, the theme of this clause would consist of the marked adjunctival 'in a bird's eye view'. But, as Downing (forthcoming) has argued, such an interpretation would suffer from not including 'my paper', since it is this element, rather than the fronted adjunct, which more properly carries the 'aboutness' of the message. In fact, 'what is being talked about' is not always co-extensive with the point of departure of the clause, and so Downing forces a separation between topic and theme, in exactly the same way as Halliday does between theme and rheme and given and new, arguing that under Halliday's 1985 formulation, theme does not always incorporate the true topic elements.

Downing's own example is as follows:

Towards the end of his life, Freud concluded
that he was not a great man,
but he had discovered great things.

According to Halliday 1985 the theme of the first clause would be 'Towards the end of his life', not including 'Freud', even though Freud is referenced in the themes of the following two clauses, and is presumably what the message is about. Downing therefore proposes, firstly, that everything up to and including the subject should be thematic (this follows Enkvist's practice (1973: 131) and, in the example above, has the effect of making the first mention of Freud thematic), and secondly, that Adjuncts should no longer be considered topical (this would mean that 'Towards the end of his life' would still be ideational, but not topical). In a slightly later paper which discusses the notion of topic more extensively (1990), Downing demonstrates what is basically the same argument, but this time applied to simple themes. In the following example,

You can define a net in one of two ways, depending on your point of view,

she claims that 'you' is not really what the clause is about, and so prefers not to call it topical, although it is still presumably ideational, perhaps of a different sub-type altogether. Different types of ideational theme will be returned to below; for the moment, however, it is the boundary between theme and rheme which is of primary concern.

In fact many analysts have chosen to extend the scope of the thematic element. On one particular occasion, Berry chooses to extend the scope of the

theme yet further, allowing in to the theme elements which can come between the subject and the main verb: 'For the purposes of this article, I have erred on the side of generosity, as it were, and included in the theme everything that anyone working in the Hallidayan tradition has ever to my knowledge advocated including. This means that I have treated as theme everything that precedes the verb of the main clause. Where a subordinate clause precedes the main clause, this too has been included' (1989: 71).

Such a definition of theme obviates the need for Halliday's 'displaced' elements in examples (1985: 64) such as:

For all his integrity and high principles, Robert pulled ...

*

'Robert', according to Halliday's definition of Theme as 'extend[ing] up to (and includ[ing]) the topical Theme' (1985: 56), is strictly no longer thematic, in spite of it being referred to as 'displaced Theme', signalled by the asterisk. This is a particularly unsatisfactory construct, and one which can be dispensed with assuming the alternative delimitation of theme.

It is fair to say that in the Systemic community there is a growing controversy over the delimitation of Theme. Berry has even suggested that it may depend upon what one is expecting to gain from the analysis. Given that there is this division of opinion, it was considered safer to adopt a multimethod approach and represent two rather different views. In the analysis of the abstracts data, therefore, the boundary between theme and rheme is drawn according to Halliday 1985, but where the two are not co-extensive, the boundary is also drawn according to Berry 1989, since she advocates a rather more 'generous' interpretation than is usual. This means that some clauses have two themes of different length, Halliday's theme being included in Berry's, as in the following example:

Abstract Set: Distinct Personality Type
Abstract: B
Clause Complex: 3

3.1 M
From the same survey - public librarians were found

top	Theme2
-----	--------

3.2
to be insecure and inadequate in social situations,
3.3
and exhibited less leadership qualities than the average

text†
university student.

In this example, for Halliday the theme of clause 3.1 is 'From the same survey', but for Berry, the theme is 'From the same survey - public librarians' (the "-†" suffix means that the 'and' is the Theme of a non-main clause; all the symbols used to code the different types of theme are defined in section 13.5.3 below).

13.5.2.2 Theme in Non-Main Clauses

Although Halliday in his thematic interpretation of the Stevenson text (1985: 64 - 67) analyses dependent as well as independent clauses (and shows the extent, but not the type, of the theme in embedded clauses), and Fries explicitly says he is working with the T-unit (see, for example, Hunt 1977: 92 - 93), the majority of analysts confine their attention to main clauses only (as does Berry 1989; she also cites Brown and Yule 1983, and Martin 1986, who follow the same practice). There are perhaps two reasons for this: firstly, it avoids some of the problems inherent in analysing dependent clauses; secondly, and more importantly, Halliday states (1985: 62) that the method of development in a text is realised principally through the thematic organisation of main clauses.

In the abstracts data, however, non-main clauses have to be analysed since there is a lot of material presented in hypotactically projected (i.e. subordinate) clauses, as in the following example:

Abstract Set: Distinct Personality Type

Abstract: D

Clause Complex: 3

3 (1 ^ +2α ^ +2"β)

3.1

The popular stereotype is examined

3.2

and the author concludes

3.3

that, in popular imagination, the librarian is seen as both diffident and severe.

Had main clauses only been analysed, the first six words of dependent clause 3.3 would not be treated as being thematic. This would be unfortunate, as Downing, for example, would argue that 'the librarian' is a topical item, is concerned with 'what the message is about', and would therefore be expected to contribute to the method of development of the text. It must be stressed that, as the preceding chapter made clear, many of the abstracts have relatively high grammatical intricacy scores, and contain substantial amounts of information realised in non-main clauses. So as not to ignore anything of potential thematic interest therefore, it was thought necessary to analyse dependent as well as independent clauses.

Unfortunately it is often difficult to draw a reliable distinction between dependent and embedded clauses, as Huddleston points out (1988: 144 - 148). Martin 1988 provides some useful comments on this point but is still unable to furnish a watertight means of differentiation. This may be because, as Matthiessen and Thompson 1988 believe, subordination is fundamentally a discourse notion, and is incapable of being satisfactorily handled purely at the level of grammar. Because certain clauses cannot be unambiguously and reliably assigned to one or other of the two categories, the decision was taken to make, in effect, *no real differentiation between the two*: both types are analysed and, together with the themes of the main clauses, each is assumed to make an equal contribution to the thematic makeup of the text. This is why 'non-main clause' is the preferred formulation to cover both dependent and embedded clauses.

While Halliday admits there are occasionally borderline cases where it is difficult to know whether to code clauses as being either embedded or dependent, he says it is important to try to distinguish between the two, arguing 'Whereas parataxis and hypotaxis are relations BETWEEN clauses..., embedding is not. Embedding is a mechanism whereby a clause or phrase

comes to function as a constituent **WITHIN** the structure of a group, which itself is a constituent of a clause' (1985: 219, original emphases). Although this appears to be a good and interpretable test, Halliday contravenes it on at least two occasions: in his analysis of 'I have quite come to a conclusion that there is nothing for me but hard work in this world' in the Stevenson text, he treats what should have been by his own criterion an embedded clause acting as a postmodifier to the headword 'conclusion' as a separate dependent clause; and in Table 7(8) he treats 'For Jack to build a house' (1985: 220) when acting as head of a nominal group (as in 'For Jack to build a house would be folly') as being embedded, when clearly it is not a constituent within the structure of the nominal group at all, rather it *is* the nominal group.

In the analysis reproduced in Appendix 13.1, instead of clause complexes containing mixtures of independent, dependent and embedded clauses (the exact mixture might be debated by different analysts), clause complexes now contain only main or non-main clauses, the basic procedure being one of maximal separation at the clause level. Now, in other words, clause complexes are quite simply split into as many separate clauses as is possible. All this has of course been made necessary by the decision to code both types of clause thematically, so no potentially interesting information is lost.

The means by which clauses are now segmented is as follows. Consider the following invented example:

Mary said she would start to drive home as soon as the film finished.

According to Berry 1975, for example, one possible interpretation of this sentence would involve a large amount of rankshifted, or embedded, items:

S	P	C		A
Mary	said		she would start to drive home	as soon as the film finished.

According to Halliday 1985, however, which on the whole prefers less embedding, the structure of the clause complex would be as follows:

$(\alpha \wedge ' \beta \alpha \wedge ' \beta x \beta)$
 Mary said ||
 she would start to drive home ||
 as soon as the film finished.

This interpretation views the clause complex as consisting of three separate clauses. In this study the tendency towards having separate clauses is taken as far as possible, with the effect that each clause contains one, and only one, lexical verb. This would mean that the hypotactically expanded verbal group 'start to drive' (1985: 255 - 260) in the above example would be split into two separate clauses, as follows:

Mary said||
she would start||
to drive home||
as soon as the film finished.

The main advantage of this approach is that clause segmentation can be performed with a greater degree of inter-analyser reliability than would have been possible with the methodology adopted in the previous chapter. Now the difficulty of deciding whether to code for constituency or dependence has been removed, and each clause, whether main or non-main, can receive a theme rheme analysis.

When they precede main clauses, according to Halliday (1985: 56 - 59), non-main clauses should be analysed twice, as in the following example:

Abstract Set: Distinct Personality Type
Abstract: E
Clause Complex: 4

4.1
Rather than adopting the 'trait' approach,

β
text

Theme of clause-complex

Theme of first clause

4.2
there should be

The double analysis shows that there are two themes at two different ranks: the whole of clause 4.1 can be treated as the theme of clause-complex 4; but only the first two words constitute the theme of clause 4.1: 'rather than' is a replative conjunctive preposition (1985: 209), and therefore textual (Table 3 (7), 1985: 54). The problem is not so much the double analysis itself, but rather how the two themes relate one to the other. Which is the more important, and, do

both signal 'aboutness', or point of departure, in exactly the same way? This seems somewhat unlikely, especially in cases such as this where the preceding non-main clause lacks representational participants. Downing (forthcoming) has gone some way toward characterising such clause-complex themes into different kinds of 'situational indicators', while Taylor (1983: 211) has noted that scientific genres tend to favour clausal themes, while historical ones prefer adjunctival themes, the former typically realising some condition, the latter spatiotemporal circumstance.

Up to now, the argument has centred on the need to analyse the themes of non-main clauses. The next problem to be discussed is somewhat different, although it is particularly evident in dependent clauses. The problem is that there are many examples such as the one above which challenge Halliday's claims that 'There is *always* an ideational element in the Theme' (1985: 53, emphasis added), and 'The Theme of *any* clause, therefore, extends up to (and includes) the topical Theme' (1985: 56, emphasis added). In Halliday's own example (1985: 63), however:

by		counting sheep	[she finally fell asleep]
struc			
Theme		Rheme	

the theme consists solely of the textual element 'by', and contains no topical element of any kind; this means that the above claims should be interpreted with regard to main clauses only. Inconsistencies of this kind are perhaps not surprising since theme is seldom discussed in anything other than main clauses. In this study on the other hand, a certain number of changes have been necessary precisely because it was felt that non-main clause material could ill afford to be overlooked.

To summarise, this section has argued that dependent clauses cannot be overlooked because, by the very nature of the abstract, there is often much of thematic importance which is realised in hypotactically projected clauses, material which ordinarily would not be analysed thematically. Such material cannot be ignored because it may well contribute to the method of development in the text. Further, since it is difficult to differentiate reliably between dependent and embedded clauses, embedded clauses must also be considered. In fact, for the purposes of this chapter, it does not matter whether non-main clauses are coded as being embedded or dependent; all that really matters is

that they both receive a thematic interpretation. For this reason the difference between embedded clauses and dependent clauses is collapsed, and they are now thought of as being simply 'non-main'. A strategy of maximal separation between clauses has been adopted, whereby each clause contains one, and only one, lexical verb. Although somewhat unusual, this was felt to be important so as to guarantee a higher degree of inter-analyser reliability.

13.5.2.3 Ellipsis

Fortunately the vast majority of clauses in the abstracts are declarative in mood, so the fact that the descriptive framework handles theme in interrogatives and imperatives altogether less well is not too serious a problem.

However there are a number of difficulties with some types of declarative clause, particularly those involving ellipsis. In the following example,

Abstract Set: Tanzania
Abstract: D
Clause Complex: 5

5.1
Tanzania has tried

top

5.2
to create model libraries

5.3
but the paper says

text	top
------	-----

5.4
this view lacks perspective,

top†

5.5
and says

text

5.6
that librarianship should be seen as theoretical and applied.

text†	top†
-------	------

there is a missing 'that' beginning clause 5.4; also presumably it is 'the paper'

which is the recoverable grammatical subject of clause 5.5. (The "-†" suffix is used to denote Themes of non-main clauses; all the symbols used to code the different types of theme are defined in section 13.5.3 below). In the next example,

Abstract Set: Tanzania
Abstract: E
Clause Complex: 11

11
11.1
The use of a "cheap manpower base" would help

top

11.2
to keep costs down
11.3
and so allow more finance for other areas of the service.

text†

the 'so' of clause 11.3 signals a resultative relationship between the clauses, there being no explicit ideational theme in the third, although it is clear that the topical element of 11.1 is intended to permeate through to 11.3. Are there, therefore, two topical themes in clause-complex 11, or only one?

Missing items could have been recovered and then analysed, but since the recovery operation is often carried out with a less than satisfying amount of inter-analyser reliability, the decision was taken in this study to code only what was actually present in the texts. This means that, in the above example, the Theme of clause 11.3 is merely 'and so'; no attempt is made to code the implicit topicality. There is therefore assumed to be only one topical theme in clause-complex 11.

13.5.2.4 Grammatical Metaphor

Data involving grammatical metaphor (Halliday 1985: chapter 10) is another source of difficulty for the framework, and one which requires fuller discussion than anyone has so far provided. The following example, from Halliday (1985: 64), constitutes a predicated theme, and, along with other types of metaphor, therefore requires a double analysis:

It was no doubt fear that he'd never get away, rather than deceit, which

int	top
-----	-----

made him mislead them.

This is an extremely confusing example for a number of reasons. Firstly, the 'which' is not coded, although its thematic status seems unquestionable: textual combined with topical, introducing the elaborating dependent clause (1985: 204). Secondly, Halliday provides one single analysis, although it seems to combine features of analysis at two different levels, the literal and the metaphorical. A properly congruent representation would have been:

It	was no doubt fear that he'd never get away, rather than deceit,	
Theme	Rheme	
top		

which	made him mislead them.	
Theme	Rheme	
text		
top		

The interpersonal flavour is better brought out on the metaphorical level, which would be represented as:

It was no doubt fear that he'd never get away, rather than deceit,
Theme
int

which	made him mislead them.	
Theme	Rheme	
text		
top		

Howeve there are problems even with this interpretation. Intuition says that there is indeed an interpersonal component to this example, but it is not present in the first analysis, and in the second, the entire first clause is coded as being interpersonally thematic, even though it includes elements which, on

a congruent level, are part of the rheme. Currently there appears to be no clear means of signalling the fact that 'no doubt' is interpersonally thematic at the metaphorical level, in spite of it occupying first position after the main verb.

There are a small number of clause complexes in the abstracts data which might be thought to involve grammatical metaphor. For example, in

Abstract Set: Distinct Personality Type
Abstract: B
Clause Complex: 10

10
10.1
On the evidence reviewed it is not possible

int	top
-----	-----

10.2
to state
10.3
that the librarian is a distinct personality type,

text†	top†
-------	------

10.4
and the usefulness of the whole psychological approach to

text	top
------	-----

librarianship is put in doubt.

top

the 'it is not possible to ...' formulation could be interpreted as an objectively oriented interpersonal metaphor of potentiality (Halliday 1985: 339 - 340). Similarly, in

Abstract Set: Distinct Personality Type
Abstract: B
Clause Complex: 6

6.1 M
With such conflicting results it seems

top	Theme2
-----	--------

6.2
that the psychological approach to librarianship has little

text†	top†
-------	------

6.3
to offer,
6.4
and it is concluded

text	top
------	-----

6.5
that it is difficult

text†	top†
-------	------

6.6
to make any sense of such findings in terms of the organisation of the library as a whole.

the 'it seems that ...' could also be interpreted as being some kind of a modalization involving probability, or some sort of interpersonal distancing device to signal the author's commitment to the following proposition. In addition, the projecting 'it is concluded ...' may be regarded by some analysts as having a metaphorical flavour.

It is difficult to see how best to analyse these alleged metaphorical usages thematically. Fortunately, such examples feature relatively infrequently in the abstracts data, and, where metaphor is involved, it was thought sufficient to provide congruent interpretations only. This is perhaps the safer option, since grammatical metaphor is still very much an emerging type of analysis.

13.5.2.5 Some Remarks on Marked Theme

The last problem to be addressed arises from the fifth hypothesis above, and concerns marked theme. The choice between marked and unmarked is discussed by Halliday in IFG (1985: 45 - 49), and perhaps slightly more fully earlier in 1967 (pages 212 - 215, and 219). In both, Halliday stresses that

thematic markedness is intimately bound up with the mood of the clause. As has already been noted, however, theme is notoriously difficult to analyse in anything but declarative clauses, so, in the interests of reliability, in this analysis it is assumed that only declarative clauses choose between the two options. Halliday says that the subject is the unmarked theme (1967: 213) in declarative clauses, and that 'The most frequent type of marked theme is the adjunct' (1967: 219). This is borne out by the data; a typical example is as follows:

Abstract Set: Vickery and Vickery, Chapter 1
Abstract: B
Clause Complex: 6

6.1 M
Recently telecommunications have overcome the distance gap

top	Theme2
-----	--------

6.2
allowing all countries
6.3
to communicate with each other.

It is important to note here that not all clauses which begin with items other than the grammatical subject are treated as being marked. For example, in 'Frankly, John is a bore', 'Frankly' is not the grammatical subject, but still would not be considered marked, since it is an interpersonal element, not an ideational one. This strictly follows Halliday's practice in chapter 3 of IFG. To sum up, then, *only declarative clauses select between marked and unmarked theme options: a theme is considered marked if and only if there is topical material in the theme which precedes the grammatical subject of the clause.*

The fact that the discussion of problems in this section ends here should not be taken to mean that no more can be said concerning the adequacy or otherwise of the framework. There are still many unresolved issues, and areas which demand further attention (notably the question of topic, and how exactly it relates to thematic organisation). For the purposes of this exploratory study, however, analysis may proceed given the provisions outlined above.

13.5.3 Summary of the Decisions Taken in the Light of these Problems

This section briefly reviews the decisions taken prior to the thematic analysis of the abstracts data, and also provides a key to the symbols used to code thematic material.

Thematic analysis is not restricted to main clauses, since a significant proportion of the material which is expected to contribute to the method of development is expressed in non-main clauses. No differentiation is made between embedded and dependent clauses, since this does not have a bearing on the theme analysis, and makes the job of clause complex segmentation considerably easier and more reliable. The fact that no differentiation is made means that clause complexes are split up further than is usual, such that each component clause contains one, and only one, lexical verb. The extent of the theme is shown in two ways: according to Halliday 1985 (everything up to and including the topical theme), and according to Berry 1989 (everything up to, but not including, the main verb). This double analysis helps to ensure that everything which might be considered part of 'what the message is about' is included within the theme.

No attempt is made to recover ellipses, and themes are given only for the congruent versions of grammatical metaphors: these last two decisions were taken to improve the internal consistency of the analysis. Lastly, this chapter takes a somewhat narrower interpretation of marked theme than is perhaps usual, so that 'Unfortunately, I couldn't go', for example, would *not* be considered marked, since 'Unfortunately' is an interpersonal element, not an ideational one.

Appendix 13.1 shows the thematic analysis of all the Information Science abstracts. In the appendix, the following symbols are used:

- **Symbols used to code clauses**

Each abstract is analysed into clause-complexes; each clause-complex consists of one or more clauses. Numbers are used to label both clause-complexes and clauses: the number before the decimal point designates the clause-complex number; the number after the decimal point designates the clause number. For example, '3.7' would be used to label the seventh clause of the third clause-complex of the abstract. When a clause has another clause 'enclosed' within it (Halliday 1985: 200), '(cont.)' denotes the uptake of the

previously mentioned enclosing clause. For example, in

Abstract Set: Distinct Personality Type
Abstract: B
Clause Complex: 2

2.1 M
In psychological studies, especially the "trait" approach, Alice Bryan (1)

top	Theme2
-----	--------

2.2
after testing 157 and 1651 female librarians,

text†

2.1 (cont.)
concluded
2.3
Librarians

the 'concluded' clause is not labelled '2.3', but rather '2.1 (cont.)', since clause 2.2 is enclosed within clause 2.1, splitting it in two parts.

• **Symbols used to code Themes**

top

denotes the topical Theme in the main clause, topical Theme being defined according to Halliday (1985: 54).

text

denotes the textual Theme in the main clause, textual Theme being defined according to Halliday (1985: 54).

int

denotes the interpersonal Theme in the main clause, interpersonal Theme being defined according to Halliday (1985: 54).

Theme2

denotes additional Thematic material, according to Berry 1989.

top†

A "-†" suffix used with any of the four symbols above denotes the Theme in non-main clauses.

Boxes split over one line are used to denote longer Themes which span more than one line of text

3.7 M

The letter 'M' after a clause number means that clause has marked Theme (Halliday 1985: 45).

β

The symbol "β" denotes Clause as Theme (Halliday 1985: 56 - 59). Such clauses are coded twice: once at the level of the clause-complex; and once at the level of the clause. β clauses would, according to Berry 1989, be treated as being part of the Theme, but in order to keep the diagrams as clear as possible however, this is not explicitly shown in the analysis.

?

Finally, a query, "?", is used to flag examples which have proved analytically problematic. A brief indication of the source of difficulty is given alongside.

13.5.4 Deriving Quantitative Measures to Test the Initial Hypotheses

This section discusses the various scores which have been allocated to the thematic analyses, while the following section, 13.5.5, will attempt a reconciliation between these scores and the judgements made of the Information Science abstracts.

The first three of the six initial hypotheses presented in section 13.4 above concern the distribution of the three types of theme Halliday calls textual, interpersonal and topical.

The success judgements concerning the Information Science abstracts are shown in Appendix 8.4.1. Each of the seventeen Information Science abstracts

has been thematically analysed following Halliday's box notation. All these analyses are shown in Appendix 13.1, but, as an example, the thematic analysis for version E of the Tanzanian set is reproduced below:

Abstract Set: Tanzania
Abstract: E

1
1.1
The article deals with some of the problems and difficulties

top

1.2
facing the Tanzanian library and documentation service.

2
2.1
It is pointed out

top

2.2
that the system of imported ideas and theories for a library service is not

text† top†

necessarily the best way of

2.3
setting up such a service.

3
3.1
The author recommends a development of libraries

top

3.2
based on the culture and social conditions of its own country.

4

4.1

Therefore it is necessary

text†	top
-------	-----

4.2

to plan for short, medium and long term projects

4.3

which may prevent future chaos or problems.

text†
top†

5

5.1

It is suggested

top

5.2

that the library service should find alternative funding methods,

text†	top†
-------	------

5.3

instead of relying solely on government funding

text†

5.4

which is often precarious.

text†
top†

6

6.1

Sensible use of such finance is also suggested

top

6.2

in order to reach a wider audience

text†

6.3

and generate further interest in literacy and libraries.

text†

7

There are several examples given.

top

8

8.1

An attempt

top

8.2

to increase literacy

8.1 (cont.)

could be helped by libraries

8.3

investing in

8.4

publishing in the native tongue rather than in a foreign language

8.5

understood by only a small minority of the population.

9

9.1

The introduction of resident authors and story times could help

top

9.2

promote general interest in literacy.

10

10.1

In order to reach a large number of people,

β

text†

10.2

it is suggested

top

10.3

that cheap multipurpose buildings are used

text†	top†
-------	------

10.4

to house libraries initially in rural areas as well as towns.

11		
11.1		
The use of a "cheap manpower base" would help		
	top	
11.2		
to keep costs down		
11.3		
and so allow more finance for other areas of the service.		
	text†	
12		
12.1		
It is suggested		
	top	
12.2		
that these are only a few ideas		
	text†	top†
12.3		
to help		
12.4		
promote the library service of Tanzania		
12.5		
and help		
12.6		
make it a more efficient and accessible service.		
13		
13.1		
It is suggested		
	top	
13.2		
that the country's underdevelopment is a challenge rather than a setback.		
	text†	top†

To test these hypotheses, the themes of each abstract are grouped together according to type and tabulated in the manner suggested by Berry 1990. The type of theme appears in the leftmost column, the number of the clause the particular theme belongs to appears in the middle column, and the themes themselves appear in the rightmost column; these are assumed to be themes of main clauses, unless suffixed by the "†" symbol, in which case they are themes of non-main clauses. Tables showing the themes of all seventeen Information Science abstracts grouped according to type are given in Appendix 13.2. One of these is reproduced overleaf as an example:

Theme Types in Tanzania, abstract version E

TYPE	THEME	
Textual	2.2	that†
	4.1	Therefore
	4.3	which†
	5.2	that†
	5.3	instead of†
	5.4	which†
	6.2	in order to†
	6.3	and†
	10.1	In order to†
	10.3	that†
	11.3	and so†
	12.2	that†
	13.2	that†
Interpersonal		
Topical	1.1	The article
	2.1	It
	2.2	the system of imported ideas and theories for a library service†
	3.1	The author
	4.1	it
	4.3	which†
	5.1	It
	5.2	the library service†
	5.4	which†
	6.1	Sensible use of such finance
	7	There
	8.1	An attempt
	9.1	The introduction of resident authors and story times
	10.2	it

Theme Types in Tanzania, abstract version E (cont.)

TYPE	THEME	
Topical	10.3	cheap multipurpose buildings†
	11.1	The use of a "cheap manpower base"
	12.1	It
	12.2	these†
	13.1	It
	13.2	the country's underdevelopment†
Clause as Theme (β)	10.1	In order to reach a large number of people†

In order to test the initial hypotheses, the number of textual and interpersonal themes have to be counted. From the table above it can be seen that abstract version E of the Tanzania set contains 13 textual themes and no interpersonal themes. Because the abstracts are of differing lengths, however, these three scores must be normalised to allow fair comparison, and so each is expressed as a percentage of the total number of clauses/non-main clauses in the abstract, and given to one decimal place. For example, Tanzania E contains 40 clauses in total, so its normalised textual theme score is $(13 \div 40) \times 100$, or 32.5%. In addition to this composite measure, separate normalised scores are calculated for theme types appearing in main and in non-main clauses; this is carried out to see if either score forms a better correlation with judges' opinions than the other.

The fourth hypothesis concerns the length of topical themes. To test this hypothesis, a score is derived for each abstract by summing the number of words occurring in its topical themes, and then dividing by the number of topical themes. This gives the average number of words per topical theme, a measure which can be fairly compared and contrasted between abstracts of different length.

Turning now to Hypothesis 5, the number of marked themes for each abstract is divided by the total number of topical themes, and expressed as a percentage. The normalisation is effected using the total number of topical themes, rather than the total number of clauses, since entry conditions for this particular system allow only topical themes to choose between marked and unmarked (see section 13.5.2.5 above).

Lastly, Hypothesis 6 concerns the distribution of clausal themes. In this case, the number of clausal themes is divided by the total number of clauses, as in the first three hypotheses.

13.5.5 Results of Testing the Initial Hypotheses

In this section each hypothesis is investigated separately (sections 13.5.5.1 to 13.5.5.6), using scores representing the overall perceived quality of the abstracts, as in previous chapters. 'Shifts away' scores are used to compare the expected and observed ranks (see Chapters 10 and 12 above). Once again, the most interesting pairings will be taken to be those which are either obviously positively correlated, or obviously negatively correlated. 'Interesting' shifts away scores are shown in bold type ('12'), uninteresting scores are

shown in normal type ('6'; see section 12.8 above).

For the Tanzania and Distinct Personality Type sets, which each consist of five abstract versions, a hypothesis will be considered to be falsified if the shifts away score is between 9 and 12 (inclusive); it will be considered to be supported if the shifts away score is between 0 and 4 (inclusive); and a shifts away score between 5 and 8 (inclusive) will be taken to mean that there is nothing of interest to be noted:

1211109	8765	43210
-ve correlation	uninteresting	+ve correlation

For the Vickery and Vickery, Chapter 1 set, which consists of four abstract versions, a hypothesis will be considered to be falsified if the shifts away score is between 6 and 8 (inclusive); it will be considered to be supported if the shifts away score is between 0 and 2 (inclusive); and a shifts away score between 3 and 5 (inclusive) will be taken to mean that there is nothing of interest to be noted:

876	543	210
-ve correlation	uninteresting	+ve correlation

For the Vickery and Vickery, Chapter 2 set, which consists of three abstract versions, a hypothesis will be considered to be falsified if the shifts away score is 4; it will be considered to be supported if the shifts away score is 0; and a shifts away score between 1 and 3 (inclusive) will be taken to mean that there is nothing of interest to be noted:

4	321	0
-ve correlation	uninteresting	+ve correlation

The results of the initial hypothesis testing are summarised in section 13.5.5.7, and in the final concluding section, 13.10.

13.5.5.1 Hypothesis 1 Results

Appendix 13.2 shows tables containing each of the abstracts' themes, grouped together according to whether they are textual, interpersonal or topical. From the data contained in this appendix, the following four tables have been derived to show the distribution of textual themes in the different abstract sets:

Distribution of Textual Themes

Abstract Set: Tanzania

Abstract Version:	A	B	C	D	E
Total number of main clauses:	12	6	7	22	13
Total number of non-main clauses:	15	7	13	18	27
Total number of clauses:	27	13	20	40	40
Number of textual Themes in main clauses:	4	2	0	11	1
Number of textual Themes in non-main clauses:	7	2	2	5	12
Total number of textual Themes:	11	4	2	16	13
Number of textual Themes in main clauses, expressed as a percentage of total number of main clauses:	33.3%	33.3%	0%	50%	7.7%
Number of textual Themes in non-main clauses, expressed as a percentage of total number of non-main clauses:	46.7%	28.6%	15.4%	27.8%	44.4%
Total number of textual Themes, expressed as a percentage of total number of clauses:	40.7%	30.8%	10%	40%	32.5%

Distribution of Textual Themes

Abstract Set: Distinct Personality Type

Abstract Version:	A	B	C	D	E
Total number of main clauses:	26	17	15	10	6
Total number of non-main clauses:	43	27	17	20	9
Total number of clauses:	73	44	32	30	15
Number of textual Themes in main clauses:	1	6	1	2	0
Number of textual Themes in non-main clauses:	9	12	8	7	6
Total number of textual Themes:	10	18	9	9	6
Number of textual Themes in main clauses, expressed as a percentage of total number of main clauses:	3.8%	35.5%	6.7%	20%	0%
Number of textual Themes in non-main clauses, expressed as a percentage of total number of non-main clauses:	20.9%	44.4%	47.1%	35%	66.7%
Total number of textual Themes, expressed as a percentage of total number of clauses:	13.7%	40.9%	28.1%	30%	40%

Distribution of Textual Themes

Abstract Set: Vickery & Vickery, Chapter 1

Abstract Version:	A	B	C	D
Total number of main clauses:	11	17	15	12
Total number of non-main clauses:	9	21	3	13
Total number of clauses:	20	38	18	25
Number of textual Themes in main clauses:	0	4	2	2
Number of textual Themes in non-main clauses:	1	10	1	4
Total number of textual Themes:	1	14	3	6
Number of textual Themes in main clauses, expressed as a percentage of total number of main clauses:	0%	23.5%	13.3%	16.7%
Number of textual Themes in non-main clauses, expressed as a percentage of total number of non-main clauses:	11.1%	47.6%	33.3%	30.8%
Total number of textual Themes, expressed as a percentage of total number of clauses:	5%	36.8%	16.7%	24%

Distribution of Textual Themes

Abstract Set: Vickery & Vickery, Chapter 2

Abstract Version:	A	B	C
Total number of main clauses:	14	13	11
Total number of non-main clauses:	14	17	7
Total number of clauses:	28	30	18
Number of textual Themes in main clauses:	3	5	1
Number of textual Themes in non-main clauses:	6	3	4
Total number of textual Themes:	9	8	5
Number of textual Themes in main clauses, expressed as a percentage of total number of main clauses:	21.4%	38.5%	9.1%
Number of textual Themes in non-main clauses, expressed as a percentage of total number of non-main clauses:	42.9%	17.6%	57.1%
Total number of textual Themes, expressed as a percentage of total number of clauses:	32.1%	26.7%	27.8%

• Results for the Tanzanian Abstracts

The first hypothesis stated that abstracts with fewer textual themes would be generally preferred over abstracts with more textual themes.

In the case of the Tanzanian abstracts, judges generally preferred version E (scoring 34 points), then version C (scoring 26), with little to separate B, A and D (scoring 21, 20 and 19 respectively). This will be referred to as the observed rank. The rank predicted by the hypothesis will be referred to as the expected rank.

Using the 'Total number of textual Themes in main clauses, expressed as a percentage of total number of main clauses' measure, the expected rank is C, E, A and B in equal third position, then D:

Abstract Version	A	B	C	D	E	Shifts away score
Preference Score	20	21	26	19	34	
% of Textual Themes in Main Clauses	33.3	33.3	0	50	7.7	
Observed Rank	E	C	B	A	D	
Expected Rank	C	E	A B		D	
						3

The shifts away score is 'interesting', so the hypothesis can be considered to be supported. However, uninteresting shifts away scores result from using both the '% of Textual Themes in Non-Main Clauses' and 'Total % of Textual Themes' measures:

Abstract Version	A	B	C	D	E	Shifts away score
Preference Score	20	21	26	19	34	
% of Textual Themes in Non-Main Clauses	46.7	28.6	15.4	27.8	44.4	
Observed Rank	E	C	B	A	D	8
Expected Rank	C	D	B	E	A	

Abstract Version	A	B	C	D	E	Shifts away score
Preference Score	20	21	26	19	34	
Total % of Textual Themes	40.7	30.8	10	40	32.5	
Observed Rank	E	C	B	A	D	
Expected Rank	C	B	E	D	A	6

In conclusion, then, only an absence of Textual Themes in main clauses can be said to be correlated with the judges' preferences.

• Results for the Distinct Personality Type Abstracts

Abstract Version	A	B	C	D	E	Shifts away score
Preference Score	8	31	22	32	27	
% of Textual Themes in Main Clauses	3.8	35.5	6.7	20	0	
Observed Rank	D	B	E	C	A	
Expected Rank	E	A	C	D	B	12

The shifts away score for Textuals in Main Clauses is 'interesting' in that it is very different from the Tanzanian set. For the Distinct Personality Type set, the hypothesis is strongly falsified: judges appear to favour a large number of Textual Themes in main clauses.

The results for non-main clauses are 'uninteresting':

Abstract Version	A	B	C	D	E	Shifts away score
Preference Score	8	31	22	32	27	
% of Textual Themes in Non-Main Clauses	20.9	44.4	47.1	35	66.7	
Observed Rank	D	B	E	C	A	
Expected Rank	A	D	B	C	E	7

However, the hypothesis is once again impressively falsified for the total percentage of Textual Themes:

Abstract Version	A	B	C	D	E	Shifts away score
Preference Score	8	31	22	32	27	
Total % of Textual Themes	13.7	40.9	28.1	30	40	
Observed Rank	D	B	E	C	A	
Expected Rank	A	C	D	E	B	
						12

• Results for Vickery and Vickery, Chapter 1

The abstracts written for the first chapter of the Vickery and Vickery book on Information Science are unusual in that this set is the only one of the four for which all three measures falsify the hypothesis:

Abstract Version	A	B	C	D	Shifts away score
Preference Score	9	23	22	26	
% of Textual Themes in Main Clauses	0	23.5	13.3	16.7	
Observed Rank	D	B	C	A	
Expected Rank	A	C	D	B	
					8

Abstract Version	A	B	C	D	Shifts away score
Preference Score	9	23	22	26	
% of Textual Themes in Non-Main Clauses	11.1	47.6	33.3	30.8	
Observed Rank	D	B	C	A	
Expected Rank	A	D	C	B	
					6

Abstract Version	A	B	C	D	Shifts away score
Preference Score	9	23	22	26	
Total % of Textual Themes	5	36.8	16.7	24	
Observed Rank	D	B	C	A	
Expected Rank	A	C	D	B	
					8

It would seem, then, that all kinds of Textual Theme are an approved strategy for the Vickery and Vickery, Chapter 1 judges.

• Results for Vickery and Vickery, Chapter 2

The scores for the three abstracts written for the second chapter of the Vickery book are interesting in that, unlike the three other sets, using one of the measures, the hypothesis is supported, and using the two others, the hypothesis is falsified. For Textual Themes in Main Clauses, the hypothesis is strongly supported:

Abstract Version	A	B	C	Shifts away score
Preference Score	17	13	18	
% of Textual Themes in Main Clauses	21.4	38.5	9.1	
Observed Rank	C	A	B	
Expected Rank	C	A	B	0

However, for Textual Themes in Non-Main Clauses and for Total Textual Themes, the hypothesis is strongly falsified:

Abstract Version	A	B	C	Shifts away score
Preference Score	17	13	18	
% of Textual Themes in Non-Main Clauses	42.9	17.6	57.1	
Observed Rank	C	A	B	
Expected Rank	B	A	C	4

Abstract Version	A	B	C	Shifts away score
Preference Score	17	13	18	
Total % of Textual Themes	32.1	26.7	27.8	
Observed Rank	C	A	B	
Expected Rank	B	A	C	
				4

In conclusion, the results differ according to the abstract set and according to which measure of Textual Theme is used. Generally, the hypothesis was falsified for the Distinct Personality Type and Vickery and Vickery, Chapter 1 sets. The hypothesis was supported for the main clause measure in the Tanzanian set. This would seem to suggest that, although they are not averse to Textual Themes in general, the Tanzanian judges prefer not to see them introducing main clauses. Similarly, the Vickery and Vickery, Chapter 2 judges prefer not to see Textual Themes introducing main clauses, but seem to like them introducing non-main clauses.

A more delicate analysis of Textual Theme type was carried out, analysing Textual Themes into continuatives, structurals and conjunctive Adjuncts (Halliday 1985: 54). The results will not be tabled, however, since the analysis revealed no interesting correlations between the sub-types of Textual Theme and the informants' preferences.

13.5.5.2 Hypothesis 2 Results

Care must also be taken in the interpretation of the interpersonal themes, since there are so few examples in the data. Their distribution is as follows:

Distribution of Interpersonal Themes

Abstract Set: Tanzania

Abstract Version:	A	B	C	D	E
Total number of main clauses:	12	6	7	22	13
Total number of non-main clauses:	15	7	13	18	27
Total number of clauses:	27	13	20	40	40
Number of interpersonal Themes in main clauses:	1	0	0	0	0
Number of interpersonal Themes in non-main clauses:	0	0	0	0	0
Total number of interpersonal Themes:	1	0	0	0	0
Number of interpersonal Themes in main clauses, expressed as a percentage of total number of main clauses:	8.3%	0%	0%	0%	0%
Number of interpersonal Themes in non-main clauses, expressed as a percentage of total number of non-main clauses:	0%	0%	0%	0%	0%
Total number of interpersonal Themes, expressed as a percentage of total number of clauses:	3.7%	0%	0%	0%	0%

Distribution of Interpersonal Themes

Abstract Set: Distinct Personality Type

Abstract Version:	A	B	C	D	E
Total number of main clauses:	26	17	15	10	6
Total number of non-main clauses:	43	27	17	20	9
Total number of clauses:	73	44	32	30	15
Number of interpersonal Themes in main clauses:	2	4	0	0	0
Number of interpersonal Themes in non-main clauses:	3	1	2	6	0
Total number of interpersonal Themes:	5	5	2	6	0
Number of interpersonal Themes in main clauses, expressed as a percentage of total number of main clauses:	7.7%	23.5%	0%	0%	0%
Number of interpersonal Themes in non-main clauses, expressed as a percentage of total number of non-main clauses:	7%	3.7%	11.8%	30%	0%
Total number of interpersonal Themes, expressed as a percentage of total number of clauses:	6.8%	11.4%	6.3%	20%	0%

Distribution of Interpersonal Themes

Abstract Set: Vickery & Vickery, Chapter 1

Abstract Version:	A	B	C	D
Total number of main clauses:	11	17	15	12
Total number of non-main clauses:	9	21	3	13
Total number of clauses:	20	38	18	25
Number of interpersonal Themes in main clauses:	0	0	0	0
Number of interpersonal Themes in non-main clauses:	0	0	0	0
Total number of interpersonal Themes:	0	0	0	0
Number of interpersonal Themes in main clauses, expressed as a percentage of total number of main clauses:	0%	0%	0%	0%
Number of interpersonal Themes in non-main clauses, expressed as a percentage of total number of non-main clauses:	0%	0%	0%	0%
Total number of interpersonal Themes, expressed as a percentage of total number of clauses:	0%	0%	0%	0%

Distribution of Interpersonal Themes

Abstract Set: Vickery & Vickery, Chapter 2

Abstract Version:	A	B	C
Total number of main clauses:	14	13	11
Total number of non-main clauses:	14	17	7
Total number of clauses:	28	30	18
Number of interpersonal Themes in main clauses:	0	0	0
Number of interpersonal Themes in non-main clauses:	1	2	1
Total number of interpersonal Themes:	1	2	1
Number of interpersonal Themes in main clauses, expressed as a percentage of total number of main clauses:	0%	0%	0%
Number of interpersonal Themes in non-main clauses, expressed as a percentage of total number of non-main clauses:	7.1%	11.8%	14.3%
Total number of interpersonal Themes, expressed as a percentage of total number of clauses:	3.6%	6.7%	5.6%

The second hypothesis stated that abstracts with fewer interpersonal themes would be generally preferred over abstracts with more interpersonal themes.

The shifts away scores are as follows:

Abstract Set	Measure	Shifts Away
Tanzania	Main	7.5
Tanzania	Non-main	6
Tanzania	Total	7.5
Distinct Personality Type	Main	8
Distinct Personality Type	Non-main	8
Distinct Personality Type	Total	12
Vickery and Vickery, Chapter 1	Main	4
Vickery and Vickery, Chapter 1	Non-main	4
Vickery and Vickery, Chapter 1	Total	4
Vickery and Vickery, Chapter 2	Main	2
Vickery and Vickery, Chapter 2	Non-main	4
Vickery and Vickery, Chapter 2	Total	2

Although there are relatively few cases in the data, there appears to be strong evidence in the Distinct Personality Type set against the hypothesis, when the total number of Interpersonals Themes is calculated. However, there appears to be strong evidence in the Vickery and Vickery, Chapter 2 set for the hypothesis, when the number of non-main Interpersonals Themes is calculated.

As above in Hypothesis 1, a more delicate sub-categorisation was carried out dividing Interpersonal Themes into i) modal Adjuncts, and ii) finite verbs and wh-interrogatives (Halliday 1985: 54). This time, the sub-categorisation does yield interesting results.

Modal Adjuncts feature only in the DPT set. The Interpersonal Themes are shown for each of the five different versions (modal Adjuncts are underlined):

Version A	8.4	How†
	14	What
	23.1	Is
	33.2	how†
	34.2	why†
Version B	5.2	what†
	5.1	<u>[Theme2: of course]</u>
	9.1	<u>Continually</u>
	9.2	often
	10.1	<u>On the evidence</u>
Version C	7.4	<u>most†</u>
	11.2	what†
Version D	1.2	is†
	2.2	whether or not†
	3.3	<u>in popular imagination†</u>
	4.4	how much fact†
	4.5	how far†
	6.4	<u>on the evidence†</u>
Version E	nil	

The reader should note that not all of these underlined items are explicitly listed in Halliday's table of modal Adjuncts (1985: 50): however, it is argued that 'Continually' can be treated as being akin to Halliday's 'usuality' Adjuncts; 'On the evidence' as a 'presumptive'; 'in popular imagination' somewhere between 'presumptive' and 'validative'; and 'most' too has an interpersonal flavour.

Bearing in mind that the observed rank for the DPT set was D, B, E, C, A, it is noticeable that modal Adjuncts appear to have found favour with the judges. This finding goes against Stainton's study (the original motivation for the hypothesis), and lends support to the idea that modal Adjuncts may be a feature of more highly regarded abstracts. Thus, although abstract writers are taught very early on not to allow their own opinions to show through in their writing, letting other people's views or positions shine through does not necessarily seem to be frowned upon.

13.5.5.3 Hypothesis 3 Results

The third hypothesis concerns the length of topical themes: abstracts which have longer topical themes are thought to be generally preferred over abstracts which have shorter topical themes. The lengths of the topical themes are charted in the following four tables:

Lengths of Topical Theme

Abstract Set: Tanzania

Abstract Version:	A	B	C	D	E
Total number of topical themes:	12	2	9	27	20
Total number of words in topical Themes occurring in main clauses:	26	0	39	55	33
Total number of words in topical Themes occurring in non-main clauses:	7	3	2	10	23
Total number of words in topical Themes occurring in either:	33	3	41	65	56
Average number of words per main topical Theme:	2.17	0	4.33	2.04	1.65
Average number of words per non-main topical Theme:	0.58	1.5	0.22	0.37	1.15
Average number of words per topical Theme:	2.75	1.5	4.55	2.41	2.8

Lengths of Topical Theme

Abstract Set: Distinct Personality Type

Abstract Version:	A	B	C	D	E
Total number of topical themes:	28	28	25	18	10
Total number of words in topical Themes occurring in main clauses:	29	67	35	23	23
Total number of words in topical Themes occurring in non-main clauses:	14	21	15	24	5
Total number of words in topical Themes occurring in either:	43	88	50	47	28
Average number of words per main topical Theme:	1.04	2.39	1.4	1.28	2.3
Average number of words per non-main topical Theme:	0.5	0.75	0.6	1.33	0.5
Average number of words per topical Theme:	1.54	3.14	2.0	2.61	2.8

Lengths of Topical Theme

Abstract Set: Vickery and Vickery, Chapter 1

Abstract Version:	A	B	C	D
Total number of topical themes:	13	24	16	18
Total number of words in topical Themes occurring in main clauses:	25	28	85	51
Total number of words in topical Themes occurring in non-main clauses:	2	25	1	8
Total number of words in topical Themes occurring in either:	27	53	86	59
Average number of words per main topical Theme:	1.92	1.17	5.31	2.83
Average number of words per non-main topical Theme:	0.15	1.04	0.06	0.44
Average number of words per topical Theme:	2.07	2.21	5.37	3.27

Lengths of Topical Theme

Abstract Set: Vickery and Vickery, Chapter 2

Abstract Version:	A	B	C
Total number of topical themes:	17	15	14
Total number of words in topical Themes occurring in main clauses:	43	31	14
Total number of words in topical Themes occurring in non-main clauses:	5	6	5
Total number of words in topical Themes occurring in either:	48	37	19
Average number of words per main topical Theme:	2.53	2.07	1.0
Average number of words per non-main topical Theme:	0.29	0.4	0.36
Average number of words per topical Theme:	2.82	2.47	1.36

An examination of the observed ranks versus those predicted by the hypothesis reveals nothing of interest in the Tanzanian abstracts:

Abstract Version	A	B	C	D	E	Shifts away score
Preference Score	20	21	26	19	34	
Average no. words per topical Theme:	2.75	1.5	4.55	2.41	2.8	
Observed Rank	E	C	B	A	D	
Expected Rank	C	E	A	D	B	6

The hypothesis is supported for the Distinct Personality Type set on the other hand:

Abstract Version	A	B	C	D	E	Shifts away score
Preference Score	8	31	22	32	27	
Average no. words per topical Theme:	1.54	3.14	2.0	2.61	2.8	
Observed Rank	D	B	E	C	A	
Expected Rank	B	E	D	C	A	4

For the Vickery and Vickery, Chapter 1 set, once again there is neither a positive nor a negative correlation, so the hypothesis is neither supported nor falsified:

Abstract Version	A	B	C	D	Shifts away score
Preference Score	9	23	22	26	
Average no. words per topical Theme:	2.07	2.21	5.37	3.27	
Observed Rank	D	B	C	A	
Expected Rank	C	D	B	A	4

The hypothesis is falsified for the Vickery and Vickery, Chapter 2 set:

Abstract Version	A	B	C	Shifts away score
Preference Score	17	13	18	
Average no. words per topical Theme:	2.82	2.47	1.36	
Observed Rank	C	A	B	
Expected Rank	A	B	C	4

In conclusion, there are differences between the sets of abstracts in terms of the explanatory power of the hypotheses. The hypothesis was supported for the Distinct Personality Type set, falsified for the Vickery and Vickery, Chapter 2 set, and nothing of interest emerged from the Tanzania set or from the Vickery and Vickery, Chapter 1 set.

13.5.5.4 Hypothesis 4 Results

Following one of Stainton's suggestions, the fourth hypothesis held that

abstracts with marked themes would be generally preferred over abstracts with fewer or no marked themes.

As was the case with the interpersonal themes (hypothesis 2 above), there are relatively few instances of marked themes in the data, so caution must be exercised in the interpretation of the numerical results. Marked themes are as follows:

Distribution of Marked Themes

Abstract Set: Tanzania

Abstract Version:	A	B	C	D	E
Total number of topical Themes:	12	2	9	27	20
Number of marked Themes:	1	0	0	2	0
Number of marked Themes, expressed as a percentage of total number of topical Themes:	8.3%	0%	0%	7.4%	0%

Distribution of Marked Themes

Abstract Set: Distinct Personality Type

Abstract Version:	A	B	C	D	E
Total number of topical Themes:	28	28	25	18	10
Number of marked Themes:	2	3	0	0	0
Number of marked Themes, expressed as a percentage of total number of topical Themes:	7.1%	10.7%	0%	0%	0%

Distribution of Marked Themes

Abstract Set: Vickery and Vickery, Chapter 1

Abstract Version:	A	B	C	D
Total number of topical Themes:	13	24	16	18
Number of marked Themes:	0	2	0	0
Number of marked Themes, expressed as a percentage of total number of topical Themes:	0%	8.3%	0%	0%

Distribution of Marked Themes

Abstract Set: Vickery and Vickery, Chapter 2

Abstract Version:	A	B	C
Total number of topical Themes:	17	15	14
Number of marked Themes:	3	0	0
Number of marked Themes, expressed as a percentage of total number of topical Themes:	17.6%	0%	0%

There are relatively few instances of marked theme in the data, so any conclusions should be treated with caution. However, it can be noted that the only interesting shifts away score occurs in the Tanzanian set, where marked themes are dispreferred, thereby falsifying the hypothesis. As was noticed in hypothesis 1, there is something of a split between the Tanzanian set and the others, as the shifts away scores demonstrate:

Tanzania:	12
Distinct Personality Type:	8
Vickery and Vickery, Chapter 1:	4
Vickery and Vickery, Chapter 2:	3

13.5.5.5 Hypothesis 5 Results

The paucity of data is also a problem in hypothesis 5, which postulated that abstracts having clausal themes would be generally preferred over abstracts

having fewer or no clausal themes. Their distribution is as follows:

Distribution of Clausal Themes

Abstract Set: Tanzania

Abstract Version:	A	B	C	D	E
Total number of clauses:	27	13	20	40	40
Number of clausal Themes (β):	0	0	0	0	1
Number of clausal Themes, expressed as a percentage of total clauses:	0%	0%	0%	0%	2.5%

Distribution of Clausal Themes

Abstract Set: Distinct Personality Type

Abstract Version:	A	B	C	D	E
Total number of clauses:	73	44	32	30	15
Number of clausal Themes (β):	0	0	0	2	1
Number of clausal Themes, expressed as a percentage of total clauses:	0%	0%	0%	6.7%	6.7%

Distribution of Clausal Themes

Abstract Set: Vickery and Vickery, Chapter 1

Abstract Version:	A	B	C	D
Total number of clauses:	20	38	18	25
Number of clausal Themes (β):	1	1	0	3
Number of clausal Themes, expressed as a percentage of total clauses:	5%	2.6%	0%	12%

Distribution of Clausal Themes

Abstract Set: Vickery and Vickery, Chapter 2

Abstract Version:	A	B	C
Total number of clauses:	28	30	18
Number of clausal Themes (β):	0	0	2
Number of clausal Themes, expressed as a percentage of total clauses:	0%	0%	11.1%

Again, there are relatively few instances of clausal theme in the data, so any conclusions should be treated with caution. However, it can be noted that, once again, the only interesting shifts away score occurs in the Tanzanian set, where clausal themes are preferred, thereby supporting the hypothesis:

Tanzania:	4
Distinct Personality Type:	5
Vickery and Vickery, Chapter 1:	4
Vickery and Vickery, Chapter 2:	1

Although the numbers involved admittedly are rather small, certain patterns in the data are evident: in every set the abstract containing the most clausal themes is also the judges' overall favourite.

This might indicate that judges sometimes appreciate having, in Downing's terms, a situational framework set up for them ready for the interpretation of the following clausal material. In the following clause complex, for example (from V&V2, version C), clause 5.1 realises what Downing terms 'a summarising initiator', signalled by the -en form of the verb:

5
5.1
Although surrounded by information
<div> <div>β</div> <div> <div>text†</div> </div> </div>
5.2
we are selective in our attention.
<div>top</div>

This has the function of 'sum[ming] up the relevant part of the preceding context in which these actions or states have been produced, at the same time providing a point of departure for something new' (Downing forthcoming).

Alternatively, the results might be explained by the fact that clausal themes break up the normal sequence of intra-clausal theme patterns, and judges may welcome this source of variety.

Without more data, however, these remarks must remain speculation.

13.5.5.6 Conclusions from the Initial Hypotheses

The results of the initial hypotheses tested above are summarised in the following table:

	Tanzania	Distinct Personality Type	V & V, Chapter 1	V & V, Chapter 2
H1a: abstracts with fewer textual Themes in their main clauses are preferred over abstracts with more textual Themes in their main clauses	✓	✗	✗	✓
H1b: abstracts with fewer textual Themes in their non-main clauses are preferred over abstracts with more textual Themes in their non-main clauses	-	-	✗	✗
H1c: abstracts with a lower total of textual Themes are preferred over abstracts with a higher total of textual Themes	-	✗	✗	✗
H2a: abstracts with fewer interpersonal Themes in their main clauses are preferred over abstracts with more interpersonal Themes in their main clauses	-	-	-	-
H2b: abstracts with fewer interpersonal Themes in their non-main clauses are preferred over abstracts with more interpersonal Themes in their non-main clauses	-	-	-	✗
H2c: abstracts with a lower total of interpersonal Themes are preferred over abstracts with a higher total of interpersonal Themes	-	✗	-	-
H3: abstracts with longer topical Themes are preferred over abstracts with shorter topical Themes	-	✓	-	✗
H4: abstracts with marked Themes are preferred over abstracts with fewer or no marked Themes	✗	-	-	-
H5: abstracts with clausal Themes are preferred over abstracts with fewer or no clausal Themes	✓	-	-	-

Hypotheses 1c, 4 and 5 are particularly interesting, because they highlight differences between the Tanzanian abstracts and the other three sets. Hypothesis 1, the claim that abstracts with fewer textual Themes would be preferred over abstracts with more textual Themes, was falsified for all abstracts except those belonging to the Tanzanian set. Hypothesis 4, the claim that abstracts with marked Themes would be preferred over abstracts with fewer or no marked Themes, was falsified only for those abstracts belonging to the Tanzanian set. Conversely, Hypothesis 5, the claim that abstracts with clausal Themes would be preferred over abstracts with fewer or no clausal Themes, was supported only for those abstracts belonging to the Tanzanian set.

The difference between the Tanzanian abstracts and those belonging to the three other sets in terms of the results of the hypothesis testing will be returned to below.

13.6 Further Hypotheses

Unfortunately, unlike textual and interpersonal themes, Halliday 1985 does not provide a more delicate characterisation of topical themes, although Drury differentiates, among other things, between human and non-human themes. Berry 1989 is more systematic and considers two types of topical theme: 'interactional' themes, which refer 'to the writer or reader(s) of the passage', and 'topic-based' or 'informational' themes, which refer to 'something which could be regarded as an aspect of the topic' (1989: 71).

Before the explanatory power of these categories can be assessed however, the definition of interactional theme needs to be modified slightly to be more appropriate to the abstracting context. Most kinds of abstract contain no overt evaluation of the source material (the exception being 'critical' abstracts which are unusual, and generally frowned upon; see Chapter 5), and so the abstractor's voice remains silent throughout. Equally, readers are seldom, if ever, referred to.

In this study, themes will be regarded as *interactional* if they refer to the author of the source material for which the abstract is a surrogate, or to other people whose work is discussed by that author, or to the reader of the text. Berry's formulation of interactional theme needs to be extended to cover the context of abstracting because more than one writer is involved. In this way,

the following topical themes from Distinct Personality Type, version D are all assumed to be *interactional*:

- 2.1 The author
- 3.2 the author
- 4.2 David Fisher
- 4.5 how far† [Theme2: the researchers†]
- 5.2 the author
- 5.4 most of the researchers†
- 6.1 David Fisher

Here 'most of the researchers' is included, although it would not be interactional according to Berry's definition (1989: 71).

In addition, themes will be regarded as *discoursal* if they refer to aspects of the source material: the complete artefact ('the article'), or any of its textual components ('the next section', 'the graph showing productivity', 'the concluding sentence'), or to studies which are themselves discussed by the source material ('these surveys', 'such studies', and so on). These are topical themes which do not refer to writer or reader, so cannot be viewed as being interactional. A particular type of informational metadiscourse (see for example Crismore 1983), discoursal themes refer not just to an aspect of the topic, but also to its textual realisation. In this way, the following topical themes from Vickery and Vickery, Chapter 1, version D are all assumed to be *discoursal*:

- 8.3 (figures on literacy growth and a table
- 9.1 Volume and variety of documentary information
- 11.1 Ranganathan's laws and any implications
- 11.2 they†

Topical themes are therefore subcategorised into three types, corresponding to the three metafunctions: discoursal themes realise those aspects of the topic which refer to the physical representation and organisation of the text on the page; interactional themes realise those aspects of the topic which refer to the writers of the text (and, by definition, to their relationship with the audience); informational themes realise those aspects of the topic which refer to the logical message communicated by the writer through the text.

These three types are all kinds of topical theme and thus they all carry ideational meaning. Interactional and discoursal themes, however, carry an additional functional component:

Informational theme:	purely ideational
Interactional theme:	ideational + interpersonal;
Discoursal theme:	ideational + textual.

Using this new categorisation, it can be hypothesised that:

• **Hypothesis 6:**

Abstracts with fewer discoursal themes will be generally preferred over abstracts with more discoursal themes;

• **Hypothesis 7:**

Abstracts with fewer interactional themes will be generally preferred over abstracts with more interactional themes; and

• **Hypothesis 8:**

Abstracts with more informational themes will be generally preferred over abstracts with more informational themes.

These hypotheses will be tested in the next section.

13.7 Testing the Further Hypotheses

Tables showing the topical themes of all seventeen Information Science abstracts grouped according to type are given in Appendix 13.3. The Tanzanian set is reproduced below as an example:

Topical Theme Types in Tanzania, abstract version A

TYPE	TOPICAL THEME	
Discoursal	9.1	Other issues
	11.1	The article
Interactional	10.1	The author
Informational	2.3	which†
	5.1	Implementation of staged development from from national to village libraries
	7.1M	because of limited funds [Theme2: maximum cost efficiency]
	7.2	methods of
	9.2	how
	10.2	fundamental librarianship principles†
	10.3	that†
	11.3	that†
	11.4	that†

Topical Theme Types in Tanzania, abstract version B

TYPE	TOPICAL THEME	
Discoursal		
Interactional		
Informational	2.4	the measures†
	4.1	that†

Topical Theme Types in Tanzania, abstract version C

TYPE	TOPICAL THEME	
Discoursal	5.1	The failure of the Hockey report (1960) and the decline in issues with the departure of the Indian and European expatriates
Interactional		
Informational	2.1	An alternative strategy
	2.3	that†
	3.1	Aspects of theoretical librarianship
	4.1	The overall objective
	5.1	The failure of the Hockey report (1960) and the decline in issues with the departure of the Indian and European expatriates
	5.3	which†
	6.1	The alternative theory
	7	Perfectionist library techniques
	8.1	Methods of

Topical Theme Types in Tanzania, abstract version D

TYPE	TOPICAL THEME	
Discoursal	1	The article
	2	The paper
	3.3	the paper
	4.1	It
	5.3	the paper
	5.4	this view†
	9.1	The paper
	10.1	the point
	13.2	the paper
Interactional		
Informational	3.1 M	Due to financial, political and moral support [Theme2: the library service]
	4.2	the library†
	4.5	options†
	5.1	Tanzania
	5.6	librarianship†
	6.1	Some principles, generalisations and assumptions
	6.3	libraries
	7.1	The lack of a cost effect approach to limited resources
	7.2	resources†
	8 M	on economics, [Theme2: the wage bill]
	9.2	tasks†
	9.3	there†
	10.2	training†
	11.1	Plans
	12.1	Small tasks
	12.3	plans
	12.5	support
	13.1	The publishing industry

Topical Theme Types in Tanzania, abstract version E

TYPE	TOPICAL THEME	
Discoursal	1.1	The article
	12.2	these†
Interactional	3.1	The author
Informational	2.1	It
	2.2	the system of imported ideas and theories for a library service†
	4.1	it
	4.3	which†
	5.1	It
	5.2	the library service†
	5.4	which†
	6.1	Sensible use of such finance
	7	There
	8.1	An attempt
	9.1	The introduction of resident authors and story times
	10.2	it
	10.3	cheap multipurpose buildings†
	11.1	The use of a "cheap manpower base"
	12.1	It
	13.1	It
	13.2	the country's underdevelopment†

It should be noted that Halliday's multiple theme concept is extended to allow 'multiple topical themes'. For example, the theme of the first clause of the third clause complex of abstract version B from the Distinct Personality Type set is:

3.1 M

From the same survey - [Theme2: public librarians]

Here the theme is treated as being part discoursal, as it refers to a study which is discussed by the source material, and part informational, as it refers to the public librarians, which can be regarded as an aspect of the topic.

Tables showing the distribution and percentages of the three types of topical theme are given on the following four pages:

Distribution of Discoursal, Interactional and Informational Themes

Abstract Set: Tanzania

Abstract Version:	A	B	C	D	E
Number of discoursal Themes, in main clauses: in non-main clauses: Total:	2 0 2	0 0 0	1 0 0	8 1 9	1 1 2
Number of interactional Themes, in main clauses: in non-main clauses: Total:	1 0 1	0 0 0	0 0 0	0 0 0	1 0 1
Number of informational Themes, in main clauses: in non-main clauses: Total:	4 5 9	0 2 2	7 2 9	11 7 18	11 6 17
Number of main topical Themes: Number of non-main topical Themes: Total number of topical Themes:	7 5 12	0 2 2	7 2 9	19 8 27	13 7 20
Number of discoursal Themes, in main clauses (%): in non-main clauses (%): Total (%):	28.6 0 16.7	0 0 0	14.3 0 0	42.1 12.5 33.3	7.7 14.3 10
Number of interactional Themes, in main clauses (%): in non-main clauses (%): Total (%):	14.3 0 8.3	0 0 0	0 0 0	0 0 0	7.7 0 5
Number of informational Themes, in main clauses (%): in non-main clauses (%): Total (%):	57.1 100 75	0 100 100	100 100 100	57.9 87.5 66.6	84.6 85.7 85
%'s are calculated by dividing by the number of main/non-main/total topical Themes. For example, the percentage of discoursal Themes in main clauses = (Number of Discoursal Themes in main clauses + Number of main topical Themes) x 100.					

Distribution of Discoursal, Interactional and Informational Themes

Abstract Set: Distinct Personality Type

Abstract Version:	A	B	C	D	E
Number of discoursal Themes, in main clauses: in non-main clauses: Total:	3 2 5	6 0 6	4 1 5	1 1 2	3 0 3
Number of interactional Themes, in main clauses: in non-main clauses: Total:	6 3 9	6 2 8	1 0 1	5 2 7	0 0 0
Number of informational Themes, in main clauses: in non-main clauses: Total:	10 6 16	8 10 18	9 10 19	4 6 10	3 4 7
Number of main topical Themes: Number of non-main topical Themes: Total number of topical Themes:	18 10 28	16 12 28	14 11 25	9 9 18	6 4 10
Number of discoursal Themes, in main clauses (%): in non-main clauses (%): Total (%):	16.7 20 17.9	37.5 0 21.4	28.6 9.1 20	11.1 11.1 11.1	50 0 30
Number of interactional Themes, in main clauses (%): in non-main clauses (%): Total (%):	33.3 30 32.1	37.5 16.7 28.6	7.1 0 4	55.5 22.2 38.9	0 0 0
Number of informational Themes, in main clauses (%): in non-main clauses (%): Total (%):	55.6 60 57.1	50 83.3 64.3	64.3 90.9 76	44.4 66.6 55.6	50 100 70
<p>%s are calculated by dividing by the number of main/non-main/total topical Themes. For example, the percentage of discoursal Themes in main clauses = (Number of Discoursal Themes in main clauses + Number of main topical Themes) x 100.</p>					

Distribution of Discoursal, Interactional and Informational Themes

Abstract Set: Vickery and Vickery, Chapter 1

Abstract Version:	A	B	C	D
Number of discoursal Themes, in main clauses: in non-main clauses: Total:	0 0 0	0 0 0	0 0 0	3 1 4
Number of interactional Themes, in main clauses: in non-main clauses: Total:	0 0 0	0 0 0	0 0 0	2 0 2
Number of informational Themes, in main clauses: in non-main clauses: Total:	11 2 13	14 10 24	15 1 16	10 4 14
Number of main topical Themes: Number of non-main topical Themes: Total number of topical Themes:	11 2 13	14 10 24	15 1 16	13 5 18
Number of discoursal Themes, in main clauses (%): in non-main clauses (%): Total (%):	0 0 0	0 0 0	0 0 0	23.1 20 22.2
Number of interactional Themes, in main clauses (%): in non-main clauses (%): Total (%):	0 0 0	0 0 0	0 0 0	15.4 0 11.1
Number of informational Themes, in main clauses (%): in non-main clauses (%): Total (%):	100 100 100	100 100 100	100 100 100	76.9 80 77.8
%s are calculated by dividing by the number of main/non-main/total topical Themes. For example, the percentage of discoursal Themes in main clauses = (Number of Discoursal Themes in main clauses + Number of main topical Themes) x 100.				

Distribution of Discoursal, Interactional and Informational Themes

Abstract Set: Vickery and Vickery, Chapter 2

Abstract Version:	A	B	C
Number of discoursal Themes, in main clauses: in non-main clauses: Total:	8 0 8	0 0 0	0 0 0
Number of interactional Themes, in main clauses: in non-main clauses: Total:	0 0 0	0 0 0	1 1 2
Number of informational Themes, in main clauses: in non-main clauses: Total:	8 4 12	11 4 15	9 3 12
Number of main topical Themes: Number of non-main topical Themes: Total number of topical Themes:	13 4 17	11 4 15	10 4 14
Number of discoursal Themes, in main clauses (%): in non-main clauses (%): Total (%):	61.5 0 47.1	0 0 0	0 0 0
Number of interactional Themes, in main clauses (%): in non-main clauses (%): Total (%):	0 0 0	0 0 0	10 25 14.3
Number of informational Themes, in main clauses (%): in non-main clauses (%): Total (%):	61.5 100 70.6	100 100 100	90 75 85.7
%s are calculated by dividing by the number of main/non-main/total topical Themes. For example, the percentage of discoursal Themes in main clauses = (Number of Discoursal Themes in main clauses + Number of main topical Themes) x 100.			

13.8 Results of Testing the Further Hypotheses

Shifts away scores were calculated for each of the three hypotheses. In fact, each hypothesis was sub-divided into three: one for main clauses; one for non-main clauses; and one for main plus non-main clauses.

As an example, calculation of the shifts away score will be demonstrated for Hypothesis 6a, which stated that abstracts with fewer discoursal Themes in their main clauses would be preferred over abstracts with more discoursal Themes in their main clauses. The percentage of discoursal Themes in main clauses (i.e. (the number of discoursal Themes in main clauses + the number of main topical Themes) x 100) for the Tanzanian set were as follows:

Abstract Version:	A	B	C	D	E
Percentage:	28.6	0	14.3	42.1	7.7

Assuming that discoursal Themes are dispreferred, these percentages yield the following expected rank:

Observed Rank:	E	C	B	A	D
(Judges' overall preferences)					
Expected Rank:	B	E	C	A	D

The shifts away score is therefore 4, as the following diagram shows:

Observed Rank:	E	C	B	A	D
Expected Rank:	B	E	C	A	D

The shifts away score of 4 therefore supports the hypothesis as there is a positive correlation between observed and expected ranks:

12	11	10	9	8	7	6	5	4	3	2	1	0
-ve correlation				uninteresting				+ve correlation				

In what follows, the shifts away scores will be given for each of the further hypotheses. As usual, 'interesting' correlations will be shown emboldened '(4)', 'uninteresting' correlations will be shown plain ('6').

13.8.1 Hypothesis 6 Results

The sixth hypothesis stated that abstracts with fewer discorsal themes would be generally preferred over abstracts with more discorsal themes.

The shifts away scores are as follows:

Tanzania:		Distinct Personality Type:	
Number of discorsal Themes,		Number of discorsal Themes,	
in main clauses (%):	4	in main clauses (%):	8
in non-main clauses (%):	8	in non-main clauses (%):	6
Total (%):	4	Total (%):	8
Vickery and Vickery, Chapter 1:		Vickery and Vickery, Chapter 2:	
Number of discorsal Themes,		Number of discorsal Themes,	
in main clauses (%):	6	in main clauses (%):	3
in non-main clauses (%):	6	in non-main clauses (%):	2
Total (%):	6	Total (%):	3

The shifts away score of 4 for Tanzania means that the hypothesis is supported: the judges of the Tanzanian set prefer abstracts with fewer discorsal themes. The Vickery and Vickery, Chapter 1 set, on the other hand, all have 'interesting' scores which falsify the hypothesis: the judges of the Vickery and Vickery, Chapter 1 set prefer abstracts with more discorsal themes. Nothing can be concluded from the Distinct Personality Type and Vickery and Vickery, Chapter 2 sets, as all the shifts away scores are 'uninteresting'; this means that the hypothesis cannot be considered to be supported or falsified.

13.8.2 Hypothesis 7 Results

The seventh hypothesis stated that abstracts with fewer interactional

themes would be generally preferred over abstracts with more interactional themes.

The shifts away scores are as follows:

Tanzania:		Distinct Personality Type:	
Number of interactional Themes,		Number of interactional Themes,	
in main clauses (%):	8	in main clauses (%):	12
in non-main clauses (%):	6	in non-main clauses (%):	8
Total (%):	8	Total (%):	10
Vickery and Vickery, Chapter 1:		Vickery and Vickery, Chapter 2:	
Number of interactional Themes,		Number of interactional Themes,	
in main clauses (%):	6	in main clauses (%):	4
in non-main clauses (%):	4	in non-main clauses (%):	4
Total (%):	6	Total (%):	4

Only two of the Tanzanian abstracts contain any interactional themes, and of these, one, version E, is the most preferred, while the other, version A, is the second least preferred. It is therefore not clear what conclusion to draw from this data; this is also apparent from the 'uninteresting' shifts away score.

In the Distinct Personality Type abstracts, however, the two most preferred abstracts, versions B and D, feature a relatively large amount of interactional themes. A's relatively high score for interactional theme should have guaranteed it a higher rank, but its nonstandard grammaticality and peculiar layout have seemingly overshadowed its interactional qualities.

In conclusion, the hypothesis is falsified in the Distinct Personality Type set, and in both Vickery and Vickery sets: thus it would appear that including interactional themes is an approved strategy for the book chapters, and for the Distinct Personality Type source material. The hypothesis was neither supported nor falsified in the case of the Tanzania set.

13.8.3 Hypothesis 8 Results

The eighth hypothesis stated that abstracts with more informational themes would be generally preferred over abstracts with more informational themes.

The shifts away scores are as follows:

Tanzania:		Distinct Personality Type:	
Number of informational Themes,		Number of informational Themes,	
in main clauses (%):	6	in main clauses (%):	12
in non-main clauses (%):	8	in non-main clauses (%):	8
Total (%):	4	Total (%):	10

Vickery and Vickery, Chapter 1:		Vickery and Vickery, Chapter 2:	
Number of informational Themes,		Number of informational Themes,	
in main clauses (%):	6	in main clauses (%):	4
in non-main clauses (%):	6	in non-main clauses (%):	4
Total (%):	6	Total (%):	4

This hypothesis is falsified for all sets except Tanzania: judges of the Distinct Personality Type, Vickery and Vickery, Chapter 1 and Vickery and Vickery, Chapter 2 sets seem to favour those abstracts with fewer informational themes. On the other hand, it appears that the judges prefer the Tanzanian abstracts to have a strong informational bias.

In conclusion, there is evidence to suggest that including other than purely informational material in the themes is an approved strategy for all but the Tanzanian abstracts. There could be two reasons for this: firstly, the abstractors may be replicating the thematic structure of the original; secondly, the abstractors may deem it necessary to include in their own abstracts some indication of the 'point' of the information, rather than simply presenting the information in a raw, ideational form.

The Tanzanian abstracts tend to represent purely ideational material in their topical themes, and that this has found favour with the judges. The other three sets, on the other hand, overlay the ideational thematic material with interpersonal information, which serves to point out the relevance of the topic to the reader and the writers' attitudes towards it, and with textual information, which serves to indicate the organisation of the text. In contrast to the Tanzanian set, this overlaying strategy has found favour with the judges of the other three sets.

It is suggested that the different choices of topical theme could be used to differentiate between informative and indicative abstracts (see section 5.3.1 above). It seems likely that an informative abstract, which 'includes at least some of the more important data, facts, observations, or conclusions presented in the original source document' (Speight 1977: 2), will tend to represent purely ideational material in its topical themes, in the attempt to represent as much

as possible of the salient information. On the other hand, it seems likely that an indicative abstract, which 'is a kind of table of contents presented in narrative style' (Gray 1970: 60), will overlay the ideational material with interpersonal and textual material, showing how the original is organised and why it may be thought to be relevant.

13.9 Discussion of the Results

The results of testing the further hypotheses are summarised in table form overleaf.

	Tanzania	Distinct Personality Type	V & V, Chapter 1	V & V, Chapter 2
H6a: abstracts with fewer discoursal Themes in their main clauses are preferred over abstracts with more discoursal Themes in their main clauses	✓	.	✗	.
H6b: abstracts with fewer discoursal Themes in their non-main clauses are preferred over abstracts with more discoursal Themes in their non-main clauses	.	.	✗	.
H6c: abstracts with a lower total of discoursal Themes are preferred over abstracts with a higher total of discoursal Themes	✓	.	✗	.
H7a: abstracts with fewer interactional Themes in their main clauses are preferred over abstracts with more interactional Themes in their main clauses	.	✗	✗	✗
H7b: abstracts with fewer interactional Themes in their non-main clauses are preferred over abstracts with more interactional Themes in their non-main clauses	.	.	.	✗
H7c: abstracts with a lower total of interactional Themes are preferred over abstracts with a higher total of interactional Themes	.	✗	✗	✗
H8a: abstracts with more informational Themes in their main clauses are preferred over abstracts with fewer informational Themes in their main clauses	.	✗	✗	✗
H8b: abstracts with more informational Themes in their non-main clauses are preferred over abstracts with fewer informational Themes in their non-main clauses	✗	✗	✗	✗
H8c: abstracts with a higher total of informational Themes are preferred over abstracts with a lower total of informational Themes	✓	✗	✗	✗

The hypotheses investigated above suggest that some kinds of topical theme may be more appropriate for some types of source text than for others.

Hypothesis 8 claimed that abstracts with more informational Themes would be preferred over abstracts with fewer informational Themes. This was supported for the Tanzanian set (H8c), but falsified for the other three. The judges seem to be indicating, then, that they deem an informational style to be more appropriate for the Tanzanian source text. Although not all of the abstract writers have chosen to avoid discursal or interactional themes, the judges' rankings imply that they are responding to the texts as if they were informative abstracts. This is in contrast to the three other sets, in which the judges' rankings imply that they are responding to the texts as if they were indicative abstracts. In the Tanzanian set the emphasis seems to be on presenting the raw information in as economical a way as possible.

Hypothesis 7 claimed that abstracts with fewer interactional Themes would be preferred over abstracts with more interactional Themes. This was falsified for every set except Tanzania. In the Distinct Personality Type set, for example, interactional themes are brought much more to the fore. There are altogether fewer topic-based themes, and the fact that informational sacrifices are made is perhaps in keeping with the source genre. The Distinct Personality Type article is, after all, a *review*, and so the abstractors have chosen their themes to give due importance not just to the information itself, but also to the people involved in its transfer.

Hypothesis 6 claimed that abstracts with fewer discursal Themes would be preferred over abstracts with more discursal Themes. This was supported for the Tanzania set (H6a and H6c) and falsified for the Vickery and Vickery, Chapter 1 set.

It is clear that a great deal hinges on the different choices of topical theme available to writers. It is suggested that different configurations of these choices appear to be one of the major linguistic motivations underlying the difference between *informative* and *indicative* abstracts, a distinction given much credence in the Information and Library Science literature (see Chapter 5 above). Informative abstracts might be thought of as being characterised by a predominance of informational themes, while indicative abstracts might be thought of as being characterised by a predominance of discursal and/or interactional themes.

It is further suggested that the three different types of topical theme serve different functions: informational themes primarily reflect the writer's desire to *enlighten*, by presenting the raw facts of the message for readers'

consideration; discorsal themes primarily reflect the writer's desire to *orient* their readers, by providing a way of navigating through the various channels in which the information is presented; interactional themes primarily reflect the writer's desire to make it easy for readers to *integrate* the knowledge, by showing readers how the information relates to the various people involved in its transfer.

13.10 Conclusions

The following conclusions are to be drawn from this chapter:

- Choice of theme is an important determinant of the perceived quality of an abstract.
- Although there are many associated problems, a Hallidayan analysis of theme was shown to be useful in discriminating between judges' preferences.
- The somewhat unusual practice of analysing the themes of non-main clauses was advocated (see 13.5.2.2), since it was argued that, in an abstract, there can be much material which is presented in hypotactically projected (i.e. subordinate) clauses. Such material may contribute to the method of development of an abstract, so it cannot afford to be ignored.
- A new type of topical theme was postulated: themes are to be regarded as *discorsal* (see 13.6 above) if they refer to aspects of the source material: the complete artefact ('the article'), or any of its textual components ('the next section', 'the graph showing productivity', 'the concluding sentence'), or to studies which are themselves discussed by the source material ('these surveys', 'such studies', and so on).
- The following two tables summarise the results of the various hypotheses tested in this chapter (ticks denote support for the hypothesis; crosses denote falsification of the hypothesis; dashes denote that no firm conclusions can be drawn):

	Tanzania	Distinct Personality Type	V & V, Chapter 1	V & V, Chapter 2
H1a: abstracts with fewer textual Themes in their main clauses are preferred over abstracts with more textual Themes in their main clauses	✓	✗	✗	✓
H1b: abstracts with fewer textual Themes in their non-main clauses are preferred over abstracts with more textual Themes in their non-main clauses	-	-	✗	✗
H1c: abstracts with a lower total of textual Themes are preferred over abstracts with a higher total of textual Themes	-	✗	✗	✗
H2a: abstracts with fewer interpersonal Themes in their main clauses are preferred over abstracts with more interpersonal Themes in their main clauses	-	-	-	-
H2b: abstracts with fewer interpersonal Themes in their non-main clauses are preferred over abstracts with more interpersonal Themes in their non-main clauses	-	-	-	✗
H2c: abstracts with a lower total of interpersonal Themes are preferred over abstracts with a higher total of interpersonal Themes	-	✗	-	-
H3: abstracts with longer topical Themes are preferred over abstracts with shorter topical Themes	-	✓	-	✗
H4: abstracts with marked Themes are preferred over abstracts with fewer or no marked Themes	✗	-	-	-
H5: abstracts with clausal Themes are preferred over abstracts with fewer or no clausal Themes	✓	-	-	-

	Tanzania	Distinct Personality Type	V & V, Chapter 1	V & V, Chapter 2
H6a: abstracts with fewer discoursal Themes in their main clauses are preferred over abstracts with more discoursal Themes in their main clauses	✓	-	✗	-
H6b: abstracts with fewer discoursal Themes in their non-main clauses are preferred over abstracts with more discoursal Themes in their non-main clauses	-	-	✗	-
H6c: abstracts with a lower total of discoursal Themes are preferred over abstracts with a higher total of discoursal Themes	✓	-	✗	-
H7a: abstracts with fewer interactional Themes in their main clauses are preferred over abstracts with more interactional Themes in their main clauses	-	✗	✗	✗
H7b: abstracts with fewer interactional Themes in their non-main clauses are preferred over abstracts with more interactional Themes in their non-main clauses	-	-	-	✗
H7c: abstracts with a lower total of interactional Themes are preferred over abstracts with a higher total of interactional Themes	-	✗	✗	✗
H8a: abstracts with more informational Themes in their main clauses are preferred over abstracts with fewer informational Themes in their main clauses	-	✗	✗	✗
H8b: abstracts with more informational Themes in their non-main clauses are preferred over abstracts with fewer informational Themes in their non-main clauses	✗	✗	✗	✗
H8c: abstracts with a higher total of informational Themes are preferred over abstracts with a lower total of informational Themes	✓	✗	✗	✗

- The hypotheses investigated above suggest that some kinds of topical theme may be more appropriate for some types of source text than for others.

- Further, it is suggested that the choice of topical theme may provide a useful means of differentiating between informative and indicative abstracts: hypotheses 6, 7 and 8 suggest that informative abstracts are characterised by a predominance of informational themes, while indicative abstracts are characterised by a predominance of non-informational themes.

Part Five

Conclusions and Suggestions

Introduction

The two final chapters of this thesis are very different from each other. Chapter 14, which describes the conclusions to be drawn from the research, is essentially a backward-looking chapter, summarising what has been achieved by the present study. Chapter 15, on the other hand, looks more towards the future, providing suggestions for further work. Unlike Chapter 14, Chapter 15 contains a good deal of new material.

The 'Conclusions' chapter has been written following R. Berry's recommendation (1986: 99) that 'A research paper should be circular in argument. That is, the formal aim of the paper should be stated in the opening paragraph; the conclusion should return to the opening, and examine the original purpose in the light of the data assembled. It is a prime error to present conclusions that are not directly related to the evidence previously presented'. Gray (1970: 42) makes a similar point, and continues 'if your report is somewhat long and fairly complex, and if you wish to present more than two or three conclusions ... list and number them consecutively ... Each conclusion ... so listed should be accompanied by the argument which links it to information in the body of the report' (1970: 44).

Bearing these various points in mind, Chapter 14 has been written to provide explicit answers to the research questions listed in section 1.5 of the opening chapter of the thesis. These questions are answered one by one, following the order established in Chapter 1. References linking each of the conclusions to specific sections in the main body of the text are given wherever possible. In the Introduction to Part Four above, the point was made that this thesis should be considered to be a 'macro pilot analysis'. The work presented here is both exploratory and enabling; that is to say, it is intended to act as a preliminary to more focused studies. Therefore, the final chapter makes some suggestions where further research might usefully be carried out.

14

Conclusions

Introduction

This chapter is unusual in that it has a generic structure unlike that of any other chapter. For reasons given in the Introduction to Part Five, the seven fundamental research questions stated in section 1.5 of the very first chapter are addressed one by one in the pages which follow.

Research Question 1:

What reasons do readers give for preferring one abstract over another?

Answer:

These are many and varied. The judges' qualitative remarks concerning the seventeen Information Science articles are presented in full in Appendix 8.4.1. Section 8.5.3.2 above provides an overview of these verbatim reports, using a classification scheme consisting of a set of thirteen codes. Each qualitative remark was given zero, one or more codes, according to the type of comment contained therein. For example, the following judge's opinion - 'poor grammar and punctuation. Too conversational' - received three codes: PUNC[TUATION]; STYL[E]; and SYNT[AX] (see Appendices 8.2 and 8.3).

Some types of comment are more common than others. The table overleaf shows the distribution of the different codes in the qualitative remarks concerning the seventeen Information Science abstracts:

Frequency of Mention	Code	Definition of Code
66	READ	The degree to which the abstract is easy or pleasant to read. Flow or coherence of writing. The degree to which the parts combine to form a recognisable and unified whole.
64	CONT	Level and appropriacy of information provided in the abstract. Amount and usefulness of detail given. Coverage. Fidelity of abstract to original. Specificity. Clarity of information itself. Accuracy.
44	LENG	The size of the abstract, as might be measured by, for example, the total number of words.
35	STRUC	The arrangement and interrelationship of parts (or the lack of them) in the abstract. Construction. Organisation of the text, for example, Introduction, Body and Conclusion. Ordering of these elements.
18	STYL	The way in which the abstract is expressed, separate from its intrinsic meaning or content, for example, formality. Tone. Appropriacy of style selected.
10	SYNT	The internal construction of sentences. Grammar.
9	LAYO	The arrangement of the material. Headings; formatting; white space. Other physical aspects of design, but excluding PARA and TYPO.
9	PARA	The division (or not) of the abstract into paragraphs; their arrangement, size, and so on.
7	TYPO	The typographic appearance of the words: the use of capitals, underlining. Abbreviations may also be coded here.
5	MISC	Rag-bag. Anything not covered by the other categories.
4	REFS	Comments concerning references or citations (or the lack of them) in the abstract.
3	PUNC	The use (or misuse) of punctuation.

The two most common types of comment concern the content (CONT) of the various abstracts and what might be termed their 'reader-friendliness' (READ). In other words, textlinguistic considerations seem to be especially important to the informants (see also Research Question 2 below).

Research Question 2:

Are all of these reasons readily interpretable? In other words, can they all be used to suggest improvements in a straightforward way, or are some less 'actionable', or, to use Enkvist's term (1978: 6) quoted in Chapter 1, are some more 'impressionistic'?

Answer:

It varies. Some of the reasons the informants have given are precise and can be acted upon immediately. Typically the more immediately 'actionable' complaints have to do with matters of layout, paragraphing, punctuation, the inclusion of references and typography. For example, it is easy to address the following two criticisms:

'uppercase irritating' (Distinct Personality Type, version A, judge 4);

'the citations tended to distract from the author's paper itself. No need for citations, I feel'
(Distinct Personality Type, version D, judge 6).

However, the judges provide a great many reasons which are more impressionistic. A selection of these appears below:

'very awkward style, clumsy ...' (Distinct Personality Type, version C, judge 7);

'incoherent' (Tanzania, version A, judge 4);

'very jerky portrayal of concepts. Didn't flow at all' (Tanzania, version A, judge 6);

'I didn't like the style of D very much - too chatty possibly' (Tanzania, version D, judge 3);

'fragmented and uncohesive' (Tanzania, version D, judge 8);

'is totally hopeless - the layout is confusing and has no real structure'
(Vickery and Vickery, Chapter 1, version A, judge 1);

'readability marred by rambling sentences'

(Vickery and Vickery, Chapter 1, version D, judge 8);

'disjointed'

(Vickery and Vickery, Chapter 2, version B, judges 3 and 4).

It is altogether harder to address this second set of criticisms, many of which are coded as 'READ', and have to do with matters of coherence and readability.

According to the researcher/analyst's intuitions, it is suggested that READ, CONT and STRUC comments can be thought to be the most impressionistic and hardest to action, LAYO, LENG, PARA, TYPO, REFS and PUNC the least impressionistic and easiest to action, with STYL and SYNT lying somewhere in the middle.

The fact that there is so much variation in the specificity of the judges' remarks is interesting. Further, it highlights a mismatch between what seems to be most important to the judges, and what seems to be most important to the writers of the abstracting guidelines literature (see Chapter 5 above). The judges make greater mention of higher level issues, but the normative literature focuses mainly on much lower level issues.

Higher level issues concern phenomena at the discourse level, such as cohesion and coherence, argument structure, readability, and so on. In a way, the fact that the judges seem to place more importance on higher level issues is very good news for the textlinguist; the phenomena they typically spend a lot of time analysing really do have a large effect on people's opinions.

Lower level issues, on the other hand, concern phenomena such as spelling, punctuation, layout, and so on. Although these issues are easy to verbalise and action (a fact which might explain why they are discussed so fully in the abstracting standards and guidelines), the judges seem to think that such concerns are relatively unimportant determinants of abstracting quality.

This research is designed to be a first step towards bridging the gap between what is considered to be important in the literature and what is considered to be important to real life users of abstracts. This study has collected opinions from people who represent genuine consumers of abstracts and has tried to make sense of these opinions by means of replicable high level linguistic analysis. In this way, people's impressionistic comments can be 'divested of needless mystique' (Enkvist 1978: 6; see section 1.2 above).

Research Question 3:

If readers do give impressionistic reasons, how should these be related to, once again in Enkvist's words (1978: 6), 'precisely describable linguistic phenomena'?

Answer:

This can be achieved by careful attempts to explain certain external measures (judges' subjective opinions) using what are hoped to be the most appropriate internal measures (certain types of linguistic analysis); the distinction between the two types of measure was introduced in section 1.2 of Chapter 1. Internal measures pertain to the linguistic attributes of text; external measures pertain to people's perceptions and expectations of text. It has been argued that external measures, which seek to describe informants' attitudes and enable texts to be compared and contrasted in terms of their differing levels of perceived quality, should precede the collection of the internal measures, which seek to describe textual phenomena. The internal measures should be used to explain the external measures, by identifying features of the discourse which give rise to the judges' preferences.

In this thesis the degree to which an internal measure can be seen to explain an external measure is assessed by attempting to falsify explicitly stated hypotheses. For the sake of illustration, one such hypothesis might be:

A sentence containing more than 400 words will be perceived to be less successful than a sentence containing less than 400 words.

Assuming a number of different sentences were tested using a number of different informants, each informant being presented with two sentences and asked which they considered more successful, two sets of measures would be taken: the first set would be internal and would consist of a set of numbers representing the word count of each sentence; the second set would be external and would consist of a set of preferences. The two sets of measures would then be reconciled by means of hypothesis testing. If the hypothesis is not falsified, then it may reasonably be concluded that the length of a sentence measured in words is a good predictor of how successful it will be perceived to be.

The above hypothesis is only an example for the sake of argument. Research Question 7 below describes the conclusions to be drawn from the actual hypotheses investigated in this thesis.

Research Question 4:

Is 'success' better explained by correlation with one, or with many, linguistic variables?

Answer:

Perceived success is better explained by correlation with many variables. The success of a text is a multivariate notion; any one linguistic measure cannot hope to explain all the variation in the judges' scores. The answer to Research Question 1 above demonstrated that consumers report many different reasons underlying their preferences. These were coded so that similar types of comment could be grouped together.

However, many of the codes are themselves extremely complex, multivariate cover terms. For example, READ, glossed as 'The degree to which the abstract is easy or pleasant to read. Flow or coherence of writing. The degree to which the parts combine to form a recognisable and unified whole', encompasses a number of, as yet, only poorly understood linguistic concepts. There is evidence in this present study to suggest that making appeal to lexical, grammatical and thematic descriptive frameworks has helped show what some of the linguistic ingredients of reader-friendly abstracts might be (see Chapters 10, 12 and 13 above, and Research Question 7 below), but two things remain unclear.

Firstly, are there still other linguistic ingredients which determine reader preferences but which have not yet surfaced? This seems intuitively quite likely, and this is one of the reasons why the work reported in this thesis should be viewed as a preliminary to a more directed investigation.

Secondly, what is the relative mix of these ingredients? This will be addressed in Research Question 5 below.

In conclusion, then, it can be noted that informants provide many different types of comment which they believe underlie their preferences, and also that many of these cannot wholly be explained by reference to any one linguistic measure. What is required, therefore, is a principled means of investigating how these different linguistic variables collectively determine the perceived quality of a text. This will be returned to in Chapter 15 below.

Research Question 5:

If success is determined by more than one linguistic variable, is it possible to say which variables are most/least important?

Answer:

Strictly speaking, the answer is no; at least not within the confines of the present study.

Some assessment of the *absolute* power of each of the internal measures can be had from consideration of the hypotheses investigated in the chapters contained in Part Four. However, it would be extremely misleading to make any strong claims about the *comparative* explanatory adequacy of the internal measures. This is because the data that has been collected for this study is, for a variety of reasons, not suited to the multivariate statistical techniques which would be required to assess the relative contribution of each of the linguistic variables to the overall level of perceived success. More complex statistical procedures will be briefly reviewed in the next chapter, the suggestion being that these would be useful in any further research concerning success in writing.

Although the properly scientific conclusion is that one cannot prioritise the variables on the basis of the available data, one can answer this research question at a more intuitive level, so long as it is made perfectly clear that the following claim should be made the subject of further hypothesis testing: grammatical intricacy and choice of Theme (Chapters 12 and 13) seem to have the greatest bearing on the judges' opinions. Judges seem to be very sensitive to the way in which clauses are combined in the abstracts, and to the different choices of Theme. This will be returned to in Research Question 7 below.

Research Question 6:

Do readers think alike? More specifically, to what extent do they agree with each other in their various preferences?

Answer:

One of the most interesting findings from this study is that judges hold widely differing views of what constitutes a successful abstract. In section 8.5.3.1 above, Kendall's Coefficient of Concordance, w , was calculated for each set of abstracts to show the extent to which judges agreed with each other's preferences. A w value of 1 signifies complete agreement between the judges, whereas a value of 0 signifies that judges are being completely random in their allocation of rank scores.

There was found to be a wide range of values, from a reasonably large amount of agreement (Distinct Personality Type, scoring 0.597), to almost complete disagreement (Vickery and Vickery, Chapter 2, scoring 0.109). The scores for all the abstracts sets are as follows:

Information Science Texts	W
Distinct Personality Type	0.597
Tanzania	0.241
Vickery and Vickery, Chapter 1	0.531
Vickery and Vickery, Chapter 2	0.109
General Knowledge Texts	W
Erosion	0.137
Cocaine	0.405
Brain Drain	0.159
Nature/Nurture	0.457

The variation in these figures lends further support to the idea of success being a multivariate concept.

Investigating success in writing is, as this thesis has tried to demonstrate, a worthwhile endeavour; however, it is always going to be hampered by two kinds of difficulty. Firstly, as has been argued, there is good reason to suppose that there are many and varied linguistic drivers of success. Secondly, the qualitative and quantitative data collected for this study suggests that judges differ over the relative importance they attach to these drivers.

There are, then, two inter-related problems which discourse theorists must face if they are to formulate a linguistically predictive model of perceived success in writing. Firstly, they must be able to specify all and only the relevant internal linguistic drivers which affect reader's perceptions of the quality of a particular type of text; it is likely that these will vary across different text-types. Secondly, they must be able to apportion different weights to each of these drivers for the different judges, in order to account for the variation in the judges' preferences. This last point will be returned to in the next chapter (see section 15.4.2 below).

Research Question 7:

Which linguistic features can help to explain readers' preferences?

Answer:

In many ways, this is the central question of this thesis because, more than any other, it addresses purely textlinguistic concerns.

A number of different types of linguistic analysis have been investigated in this thesis; the reasons why these in particular were chosen are documented in the Introduction to Part Four.

Broadly, the internal measures which best explain readers' preferences have been derived from analyses of lexical density and lexical variation (Chapter 10), from analysis of grammatical intricacy (Chapter 12) and from analysis of Themes (Chapter 13).

The two other types of analysis which were piloted (Generic Structure in Chapter 9 and Cohesion in Chapter 11) yielded no significant results, mainly because many of the decisions made during analysis were, in the technical sense, *unreliable*. External measures concern subjective opinion and it is quite proper that people's intuitions are made the subject of study. However, there is a sense in which the internal measures should be as objective as possible, thereby maximising reliability and validity. The pilot studies of generic structure and of cohesion were less successful primarily due to the fact that too much subjectivism crept into the analyses; too much appeal was made to the analyst's intuitions, intuitions which could not be guaranteed to be shared by other researchers. These types of analysis are not yet sufficiently reliable to allow rigorous and credible hypothesis testing. The reader is referred to the concluding sections of each chapter for further details.

The analyses which were taken further (Chapters 10, 12 and 13), on the other hand, are altogether more reliable. The results from each of these chapters will be briefly summarised below.

Lexical Density and Lexical Variation

The following hypotheses were tested in Chapter 10 (see section 10.3):

- Hypothesis 1 stated that abstracts with higher lexical density scores would be generally preferred over abstracts with lower lexical density scores.

- Hypothesis 2 stated that abstracts with higher lexical variation scores would be generally preferred over abstracts with lower lexical variation scores.

These hypotheses were tested because one would expect good quality abstracts to contain large amounts of new information. In fact, both these hypotheses were falsified: Hypothesis 1 was falsified for all four sets; Hypothesis 2 was falsified for all sets except Vickery and Vickery, Chapter 2.

In the case of lexical density especially, precisely the reverse seemed to hold: counter-intuitively, the better abstracts were characterised by lower levels of lexical density. A correspondence was noted between the lexical density scores and a number of the judges' qualitative comments. Abstracts with lower lexical density scores tend to be described as being 'clear'. It is tentatively suggested that low levels of lexical density and lexical variation are more the mark of 'reader-friendly' abstract writing, whereas higher levels of lexical density and lexical variation characterise abstracts which contain more information but are correspondingly harder to process.

Grammatical Intricacy

Chapter 12 above argued that grammatical intricacy can be measured in two ways: the first calculates the *extent* to which clauses combine in a text to form larger structural units; the second calculates a text's reliance on each of the different *ways* in which clauses can be combined. With regard to the first of these measures:

- Hypothesis 1 (see section 12.5) stated that abstracts with a larger amount of clause level complexity would be generally preferred over abstracts with a smaller amount of clause level complexity.

The hypothesis was supported for the Distinct Personality Type abstracts and for the Vickery and Vickery, Chapter 2 abstracts. Further, there was some slight support for the hypothesis in the case of the Chapter 1 abstracts. Also, a weak correlation was noted between the ranks in the Tanzanian data. It would seem then that clause level complexity is a reasonably good predictor of the judges' overall preferences.

Further investigations were carried out concerning the relationship between readers' opinions and the different ways in which clauses can be combined. It was found that some clause combining strategies were noticeably preferred by the judges, while others were noticeably dispreferred. However, these preferences were not shared across the different abstract sets. The following summary table is reproduced from section 12.10 above (ticks denote support for the hypothesis; crosses denote falsification of the hypothesis; dashes denote that no firm conclusions can be drawn):

	Tanzania	Distinct Personality Type	V & V, Chapter 1	V & V, Chapter 2
Paratactic Elab	✕	✕	✓	-
Paratactic Ext	✕	✓	-	✕
Paratactic Enh	-	-	-	-
Paratactic Proj	-	✓	-	-
Hypotactic Elab	✕	-	✓	-
Hypotactic Ext	-	✕	-	-
Hypotactic Enh	-	✓	-	-
Hypotactic Proj	-	-	✓	-
Total Parataxis	✕	✓	-	-
Total Hypotaxis	✓	-	✓	-
Total Elab	✕	✕	✓	✓
Total Ext	✕	✓	-	✕
Total Enh	-	✓	-	-
Total Projection	-	✓	✓	-

For example, there is a positive correlation between the total amount of projection and the judges' ranks for both the Distinct Personality Type and the Vickery & Vickery, Chapter 1 abstracts. This would suggest that the judges respond well to abstracts which feature a relatively large amount of projected material. However, projection in both the Tanzania and the Chapter 2

abstracts is not a good predictor either of success or unsucccess.

It is tentatively suggested that the distribution of clause relations may provide a useful way of differentiating between sub-types of abstract. One working hypothesis might be that indicative abstracts feature relatively large amounts of projection and relatively small amounts of enhancement, while informative abstracts feature relatively large amounts of enhancement and relatively small amounts of projection.

Theme Choices

Judges seem to be particularly sensitive to choice of theme. This is evident from their qualitative pronouncements and from their quantitative rankings. Although relatively difficult to analyse, the study of theme was rewarding since it was found to have a large bearing on the judges' opinions.

A new type of theme was identified to complement the two already existing sub-types of topical theme, interactional and informational (see Berry 1989: 71). Themes are to be regarded as *discoursal* (see 13.6 above) if they refer to aspects of the source material, or to studies which are themselves discussed by the source material.

Altogether eight hypotheses were investigated in Chapter 13. The results were as follows (ticks denote support for the hypothesis; crosses denote falsification of the hypothesis; dashes denote that no firm conclusions can be drawn):

	Tanzania	Distinct Personality Type	V & V, Chapter 1	V & V, Chapter 2
H1a: abstracts with fewer textual Themes in their main clauses are preferred over abstracts with more textual Themes in their main clauses	✓	✗	✗	✓
H1b: abstracts with fewer textual Themes in their non-main clauses are preferred over abstracts with more textual Themes in their non-main clauses	-	-	✗	✗
H1c: abstracts with a lower total of textual Themes are preferred over abstracts with a higher total of textual Themes	-	✗	✗	✗
H2a: abstracts with fewer interpersonal Themes in their main clauses are preferred over abstracts with more interpersonal Themes in their main clauses	-	-	-	-
H2b: abstracts with fewer interpersonal Themes in their non-main clauses are preferred over abstracts with more interpersonal Themes in their non-main clauses	-	-	-	✗
H2c: abstracts with a lower total of interpersonal Themes are preferred over abstracts with a higher total of interpersonal Themes	-	✗	-	-
H3: abstracts with longer topical Themes are preferred over abstracts with shorter topical Themes	-	✓	-	✗
H4: abstracts with marked Themes are preferred over abstracts with fewer or no marked Themes	✗	-	-	-
H5: abstracts with clausal Themes are preferred over abstracts with fewer or no clausal Themes	✓	-	-	-

	Tanzania	Distinct Personality Type	V & V, Chapter 1	V & V, Chapter 2
H6a: abstracts with fewer discoursal Themes in their main clauses are preferred over abstracts with more discoursal Themes in their main clauses	√	.	x	.
H6b: abstracts with fewer discoursal Themes in their non-main clauses are preferred over abstracts with more discoursal Themes in their non-main clauses	-	.	x	.
H6c: abstracts with a lower total of discoursal Themes are preferred over abstracts with a higher total of discoursal Themes	√	.	x	.
H7a: abstracts with fewer interactional Themes in their main clauses are preferred over abstracts with more interactional Themes in their main clauses	-	x	x	x
H7b: abstracts with fewer interactional Themes in their non-main clauses are preferred over abstracts with more interactional Themes in their non-main clauses	-	.	.	x
H7c: abstracts with a lower total of interactional Themes are preferred over abstracts with a higher total of interactional Themes	-	x	x	x
H8a: abstracts with more informational Themes in their main clauses are preferred over abstracts with fewer informational Themes in their main clauses	-	x	x	x
H8b: abstracts with more informational Themes in their non-main clauses are preferred over abstracts with fewer informational Themes in their non-main clauses	x	x	x	x
H8c: abstracts with a higher total of informational Themes are preferred over abstracts with a lower total of informational Themes	√	x	x	x

Hypotheses 1c, 4 and 5 are interesting because they highlight differences between the Tanzanian abstracts and the other three sets.

Hypothesis 1, the claim that abstracts with fewer textual Themes would be preferred over abstracts with more textual Themes, was falsified for all abstracts except those belonging to the Tanzanian set.

Hypothesis 4, the claim that abstracts with marked Themes would be preferred over abstracts with fewer or no marked Themes, was falsified only for those abstracts belonging to the Tanzanian set.

Hypothesis 5, the claim that abstracts with clausal Themes would be preferred over abstracts with fewer or no clausal Themes, was supported only for those abstracts belonging to the Tanzanian set.

Hypothesis 6 claimed that abstracts with fewer discoursal Themes would be preferred over abstracts with more discoursal Themes. This was supported for the Tanzania set (H6a and H6c) and falsified for the Vickery and Vickery, Chapter 1 set.

Hypothesis 7 claimed that abstracts with fewer interactional Themes would be preferred over abstracts with more interactional Themes. This was falsified for every set except Tanzania. In the Distinct Personality Type set, for example, interactional themes are brought much more to the fore. There are altogether fewer topic-based themes, and the fact that informational sacrifices are made is perhaps in keeping with the source genre. The Distinct Personality Type article is, after all, a *review*, and so the abstractors have chosen their themes to give due importance not just to the information itself, but also to the people involved in its transfer.

Hypothesis 8 claimed that abstracts with more informational Themes would be preferred over abstracts with fewer informational Themes. This was supported for the Tanzanian set (H8c), but falsified for the other three.

The judges seem to be indicating, then, that they deem an informational style to be more appropriate for the Tanzanian source text. Although not all of the abstract writers have chosen to avoid discoursal or interactional themes, the judges' rankings imply that they are responding to the texts as if they were informative abstracts. This is in contrast to the three other sets, in which the judges' rankings imply that they are responding to the texts as if they were indicative abstracts. In the Tanzanian set the emphasis seems to be on presenting the raw information in as economical a way as possible.

It is clear that a great deal hinges on the different choices of topical theme available to writers. It is suggested that different configurations of these choices appear to be one of the major linguistic motivations underlying the difference between *informative* and *indicative* abstracts, a distinction given

much credence in the Information and Library Science literature (see Chapter 5 above). Informative abstracts might be thought of as being characterised by a predominance of informational themes, while indicative abstracts might be thought of as being characterised by a predominance of discoursal and/or interactional themes.

It is further suggested that the three different types of topical theme serve different functions: informational themes primarily reflect the writer's desire to *enlighten*, by presenting the raw facts of the message for readers' consideration; discoursal themes primarily reflect the writer's desire to *orient* their readers, by providing a way of navigating through the various channels in which the information is presented; interactional themes primarily reflect the writer's desire to make it easy for readers to *integrate* the knowledge, by showing readers how the information relates to the various people involved in its transfer.

In conclusion, this thesis has advocated adoption of a multimethod design, or the combining of techniques with complementary strengths and non-overlapping weaknesses. Data has been collected and analysed in a variety of different ways to minimise the inaccuracies which inevitably creep into studies which attempt to chart people's attitudes and behaviour.

Further, this thesis has tried to demonstrate that success in writing, a concept that is central to discourse theory as defined by Morgan and Sellner (1980: 196), can and should only be investigated by seeking out and valuing the opinions of readers, even though their subjective remarks may be very impressionistic. Further, such external measures should drive both the selection and the application of internal measures, or the linguistic frameworks which the textlinguist adopts to explain the readers' preferences.

The success of discourse theory crucially relies upon the tools and techniques which are used to reconcile these two types of measure. Chiefly, these concern the rigorous testing of explicitly stated hypotheses and a means of predicting which particular variables hold the greatest predictive power (statistical techniques will be discussed in Chapter 15 below).

15

Suggestions for Further Research

15.1 Introduction

This chapter is divided into three sections: section 15.2 suggests how further, more sensitive, external measures could be collected to reflect a more representative spread of subjective opinion; section 15.3 suggests a number of additional internal measures which might usefully be taken from the texts; and section 15.4 sketches a methodology to enable a principled and trustworthy reconciliation between the external and the internal measures, between the informants' judgements and the linguistic features present in the texts in other words.

15.2 Collecting Further External Measures

Given the time that was available to collect the external measures on which this study was based, the data collection phase was very successful and served the aims and objectives of this thesis very well. However, no attempt to collect data can ever be perfect and the work described in Part Three above is no exception. Specifically, the external measures described in this thesis have the following limitations:

1) Although the qualitative data indicated that the reasons underlying the judges' preferences were many and varied (Chapter 8), there is no guarantee that they exhaust all the different possible drivers of success as perceived by the reader. There may have been some drivers which, for whatever reason, did not emerge from the judges' responses.

2) Further, because the consumers' questionnaires were designed to collect judges' reasons in an unstructured way, there is no means of knowing how

each judge feels about all the different drivers elicited.

3) Also, there was a tacit assumption in Chapter 8 above that those reasons verbalised most frequently (READ, CONT and STRUC comments) were in fact the most important determinants of readers' preferences. This is not necessarily the case.

4) Further, because of the exploratory nature of the present study with its emphasis on qualitative rather than quantitative issues, it has not been possible within the confines of this research to show i) how these different drivers of readers' perceptions inter-relate, and ii) which are the best predictors of the judges' overall rankings.

5) Because the questionnaires instructed the consumer informants to rank, as opposed to score, the different abstracts, the resultant data can only indicate which abstracts are better than others; no strong claims can be made concerning their absolute quality.

The following suggestions are intended to show how these shortcomings could be addressed in further work. This section describes further means of data collection which, together with more sophisticated statistical analysis (see section 15.4 below), could help to overcome the weaknesses in the present work. Section 15.2.1 describes one means of eliciting more representative drivers of perceived success principally through the use of focus groups. Section 15.2.2 outlines a proposal for a somewhat different type of questionnaire from those described in Chapter 7 above, which, it is argued, could help explicate the underlying relationships between different sets of internal and external measures in a more rigorous fashion than has been possible within the constraints of this thesis.

15.2.1 Collecting Further External Qualitative Data: the Grocery List

Qualitative data can be obtained using in-depth interviews, personal narratives, ethnographies, questionnaires (the preferred method for this study; see Chapter 7) and so on. However, the most common means of collecting qualitative data in marketing circles is the focus group (Templeton

1987: 3). Focus groups usually consist of between 5 and 10 informants specially selected to represent the types of opinion marketers are interested in. These people talk openly in a small group with the researcher or 'moderator' present, the function of the moderator being to steer the conversation very gently according to a general discussion outline agreed with the participants in advance.

For present purposes, the main advantage of the focus group is that the resultant opinions are relatively 'natural', they are expressed in the participants' own words, and they are *more representative of the collective feeling of the group*. One of the distinguishing characteristics of the focus group is that participants can 'bounce ideas off one another' and opinions may emerge which otherwise would not, thanks to the principle that 'most people feel more comfortable talking about almost any subject when they are involved in a discussion as part of a group' (Greenbaum 1988: 18).

It is suggested that focus groups could usefully be adopted to act as the primary means of eliciting 'the grocery list' of features that may affect readers' perceptions of the degree of success exhibited by a text. This grocery list should consist of each and every feature of a text which the focus group informants report as having either a positive or negative effect on its overall perceived quality. It is further suggested that each item on the grocery list should be presented to larger sets of different respondents for quantitative evaluation. This should ensure that the drivers of success thought to be important by one specific group are shared by larger sections of the population.

To sum up, then, it is recommended:

- that focus groups be considered as a means of eliciting all the relevant drivers of success as perceived by the reader, then
- that questionnaires be used to quantify a large sample of people's opinions concerning those drivers.

Questionnaires to realise the second of these recommendations will be discussed in the next section.

15.2.2 Collecting Further External Quantitative Data

This section discusses a particular kind of questionnaire which is increasingly being used by industries to assess the degree to which customers are satisfied with the services they are offered. The instrument is called 'ServQual', standing for Service Quality, and has been developed by American marketers and psychologists (Zeithaml et al 1990). It is submitted that many of these ideas could profitably be imported into discourse theory.

The basic premise underlying their model is that customer satisfaction (a concept which cannot, they argue, be measured directly in a reliable and valid way) can be quantified as the gap between customers' perceptions and their expectations of the service they are given (two concepts which can be measured directly). The discourse theory equivalent of this premise would be that perceived success can be quantified as the difference between readers' expectations and their perceptions of a text.

Thus perceived quality could be assessed by a questionnaire such as the following:

	Much better than I expected	Better than I expected	About what I expected	Worse than I expected	Much worse than I expected
1. Compared to what it should be, the readability of the text was ...					
2. Compared to what it should be, the level of information given was ...					
3. Compared to what it should be, the tightness of the argument was ...					
etc.					
etc.					

There are a number of points to note here. Firstly, respondents would be invited to tick the box which best represents their views towards each of the numbered attitude statements. Secondly, each of these statements shown in bold should realise one of the items on the grocery list collected from the focus groups. This means that there will be a large number of questions (15 - 20 perhaps) asked of each respondent. Thirdly, the 'compared to what it should be' formulation encapsulates the idea of readers' expectations being matched (or not) by their perceptions.

The advantage of this questionnaire over the ones described in Chapter 7 above is that respondents are asked to evaluate the text in terms of each of the relevant component drivers of perceived success. These more sensitive external measures would enable closer couplings to be made with the internal measures, the linguistic features of the text, in other words. Also, predictions concerning the relative importance of these drivers, and how exactly these collectively determine the overall perceived success of the text can then be made using standard multivariate techniques (see section 15.4 below).

15.3 Collecting Further Internal Measures

There are three types of additional internal measure which could be collected in further research.

Firstly, those descriptive frameworks which were discussed in the Introduction to Part Four above but which were not taken further in this study could be investigated in the future. These include analyses of Transitivity, Groups and Phrases, and Grammatical Metaphor. Given the encouraging results reported in Chapter 13 above concerning informational versus non-informational topical Theme, an analysis of metadiscourse would seem to be an avenue well worth exploring.

Secondly, there may well be other types of analysis not discussed anywhere in this thesis which could be used to explain informants' preferences. A focus group of textlinguists could be held to identify promising types of analysis not considered so far.

Thirdly, the three types of internal measure investigated in Part Four above - lexical texture, grammatical intricacy and choice of Theme - could be investigated in greater detail. Some specific suggestions are made in the following three sections.

• Suggestions for Further Research on Lexical Texture

1) In Chapter 10, it was pointed out that one must be very careful when comparing and contrasting lexical variation scores for texts of differing length. Normalising for length by examining the first n words of a set of texts is probably the best general strategy (Biber 1988: 239), but in smaller texts such as abstracts, curtailing texts too drastically may interfere with their perceived generic structure; the resultant lack of structural unity may bias informants' responses adversely. One way round this problem would be to collect abstracts from journals which have an editorial policy stipulating a maximum word length.

2) In this study single lexical density and lexical variation scores were computed for each of the abstracts in their entirety. In further work it would be interesting to investigate scores for individual clause complexes, and compute standard deviation scores. It could be hypothesised that judges may prefer abstracts with a greater variety of levels of lexical density and variation.

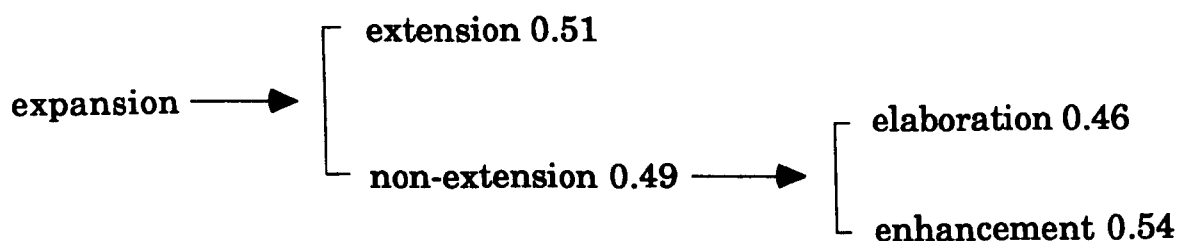
3) The approach taken in this thesis concerning the relationship between lexical density and grammatical intricacy is somewhat different from the stance taken by Halliday. For Halliday, there exists a reciprocal relationship between the two. This study treats them as if they were independent variables. In Chapter 10, no appeal to grammatical complexity is made in the calculation of lexical density. Similarly, in Chapter 12, no appeal to lexical complexity is made in the calculation of grammatical intricacy. This was done so that the effects of one could be investigated without fear of bias from the other. However, it seems clear that more work needs to be carried out to investigate how exactly these two concepts interrelate, preferably in a variety of different text-types.

4) In a study comparing and contrasting spoken and written language, Biber (1988: 104 - 105) has noticed that texts which have high informational focus and are carefully integrated (texts which lie further towards the written end of the mode continuum, in other words) are characterised by: a high type/token ratio (high lexical variation); have high frequencies of nouns, prepositional phrases and attributive adjectives; and contain longer words. It is therefore suggested that an analysis of groups and phrases following Chapter 6 of Halliday 1985 may further help to explain some of the judges' reactions concerning 'chatty' and 'verbose' language.

• **Suggestions for Further Research on Grammatical Intricacy**

The analysis of grammatical intricacy described in Chapter 12 above is clearly at a very early stage; there is much to be discovered concerning the way clause combining relations affect the perceived success of a text. It is somewhat surprising that more investigations have not been carried out in this area within Systemic-Functional Linguistics, especially since there has been so much work conducted by researchers from different schools (see section 12.2 above). A Systemic-Functional approach does seem to hold great promise and the logico-semantic apparatus Halliday proposes is new and distinctively systemic.

Nesbitt and Plum adopt a probabilistic approach to clause combining and motivate a slightly different system network for expansion (1988: 26) from the one which is implicit in Halliday 1985:



This alternative interpretation is motivated by the relative frequencies of the different options observed in a total of 123 texts.

Although this is not explicitly argued above in Chapter 12, according to the present researcher's intuitions, this may well map onto the distinction between informative and indicative abstracts. It is tentatively suggested that extension, particularly when realised paratactically, is more associated with indicative than with informative abstracts. Further, it is tentatively suggested that elaboration and enhancement are more associated with informative than with indicative abstracts. This could be because extended clauses tend to realise *new* types of information, whereas elaborated and enhanced clauses tell the reader more about the *given* information. The suggestion is that indicative abstracts, if they are to cover a reasonable amount of the source material, must constantly move on to new material, whereas informative abstracts, which may home in on what are considered to be the most important aspects of the source text, may spend longer on a certain topic, 'restating in other words, specifying in greater detail, commenting or exemplifying' (Halliday's functional definition of elaboration), or 'qualifying it with some circumstantial feature of time, place, cause or condition' (Halliday's functional definition of elaboration, 1985: 196 - 197). However, these remarks are very speculative and require a great deal of further research.

The similarities and differences between indicative and informative abstracts really need to be investigated more systematically than was possible in this study. Although the students were prompted to write informative abstracts, clearly some have not followed the advice the lecturers offered them. In some ways this made the present study more interesting. However, in future work, it would be instructive to carry out more controlled data collection such that judges' responses to different sub-types of abstract could be compared and contrasted in a more rigorous fashion.

• **Suggestions for Further Research on Themes**

The Hallidayan approach to Theme was largely followed in Chapter 13 above. This could usefully be augmented by attempting to chart what Fries calls 'method of development' (Fries 1983), although it is not yet clear how this could be done in a sufficiently reliable way to enable precise hypothesis testing. 'Flow' reappears in the judges' qualitative opinions, and while some attempt has been to explain these comments in terms of cohesion and generic structure, there remains much work to be done to relate this impressionistic term to precise linguistic phenomena. The reader is referred to Fries (forthcoming).

The Prague School notion of Thematic Progression (see for example Danes 1974) could be investigated in further work on abstracts and abstracting; Nwogu 1990 was able to report encouraging results using TP in his study of the differences between journalistic and more academic treatments of the same medical subject matter. Distinct Personality Type version C, the second least preferred abstract, according to the analyst's intuitions, starts promisingly enough but then lapses into repeated (or 'constant' in Prague School terminology) themes. Given that space is at a premium, constant themes run the risk of clogging an abstract with informationally redundant material and so it could be hypothesised that abstracts with linear or derived themes will be preferred over abstracts with constant themes. Alternatively, it could be hypothesised that linear and derived theme patterns are more associated with informative than with indicative abstracts, whereas a constant thematic pattern is more associated with indicative than with informative abstracts.

The treatment of Theme in Chapter 13 pointed out some interesting correspondences between the judges' preferences and the different thematic choices made by the abstract writers. However, no account was taken of the *positioning* of the different types of Theme in the abstracts. The following diagram charts the different types of topical Theme in Distinct Personality Type, version D:

Topical Theme Types in Distinct Personality Type, version D

Clause Number	Discoursal	Interactional	Informational
1.1	The article	The author	is†
1.2			whether or not† [
2.1			Theme2: psychology†]
2.2	the author	David Fisher	The popular stereotype
3.1			the librarian†
3.2			how much fact† [
3.3	David Fisher	the author	Theme2: there†]
4.2			how far† [Theme2: the researchers†]
4.4			the word†
4.5	the word†	most of the researchers†	"it
5.2			the librarian†
5.3			"The utility of the whole
5.4	David Fisher	David Fisher	psychological approach
6.1			There
6.6			
6.8	David Fisher	David Fisher	
7			
8			

This diagram makes clear just how much switching there is between different kinds of topical Theme. Further, the switching is quite regular; no one kind of topical Theme is maintained for longer than three clauses. Distinct Personality Type version D is the most preferred abstract overall, so perhaps the judges like a more dynamic kind of topicality. A rough measure of this switching can be had by counting the number of lines drawn between different topical Theme types (11 in the above diagram) and then dividing by the total number of clauses (18). It could be hypothesised that abstracts with more topical Theme switches will be preferred over abstracts with fewer or no topical Theme switches.

Berry 1989 and forthcoming) also seems to be a promising area for further exploration.

15.4 Reconciling the Two Types of Measure

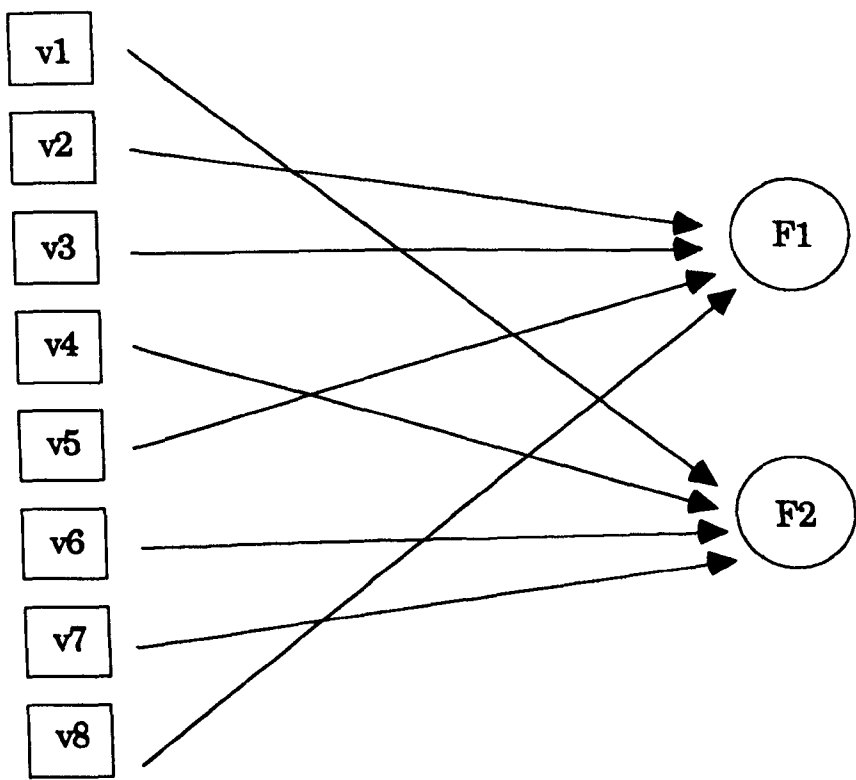
This section details the kinds of statistical analysis which could be performed on the type of questionnaire sketched above in section 15.2.2. Two techniques will be discussed in this section: Factor Analysis (section 15.4.1), which is basically a means of reducing large bodies of data to more manageable size; and Discriminant Function Analysis (a particular kind of regression, section 15.4.2), which can be used to categorise cases into groups, order variables according to importance, and so on. An excellent introduction to these and many other techniques for collecting and analysing data which could be very useful for the discourse theorist is Churchill 1991.

As a macro-pilot analysis (see the Introduction to Part Four), the purpose of the present study was to identify which of the many different types of linguistic analysis which could have been used to describe the data offered the best chance of explaining the most interesting of the informants' judgements. Ideally it would have been instructive to test a large number of potentially useful linguistic tools, but, for practical reasons, five candidate analyses only (Chapters 9 - 13) were pilot tested.

Three of these were taken further, and made the subject of a larger scale investigation (Chapters 10, 12 and 13). Such 'homing in' on particularly interesting-looking phenomena is a natural process for textlinguists, who are typically confronted with massive amounts of complex data and seemingly endless means of describing them, all of them problematic. However there remains the problem of knowing at the end of the research enquiry how much of the subjective data has been satisfactorily explained by any one particular micro-analysis. For the discourse theorist interested in what constitutes successful writing, this problem is of fundamental importance. Clearly the concept of success is multifaceted (Chapter 14), and, even for one well-defined genre, it is not yet possible to specify how many different factors or dimensions are involved, let alone which are the most important. Given these uncertainties, it is particularly important to attempt to assess how trustworthy the 'results' are. One means of doing this would be to employ multivariate statistical techniques.

15.4.1 Factor Analysis

Factor Analysis seeks to reduce a large number of variables to a smaller set of underlying, and uncorrelated, 'factors'. In other words, it summarises the data, and can help to show the textlinguist which features are the principal markers of difference between the texts. This technique can be represented diagrammatically as follows:

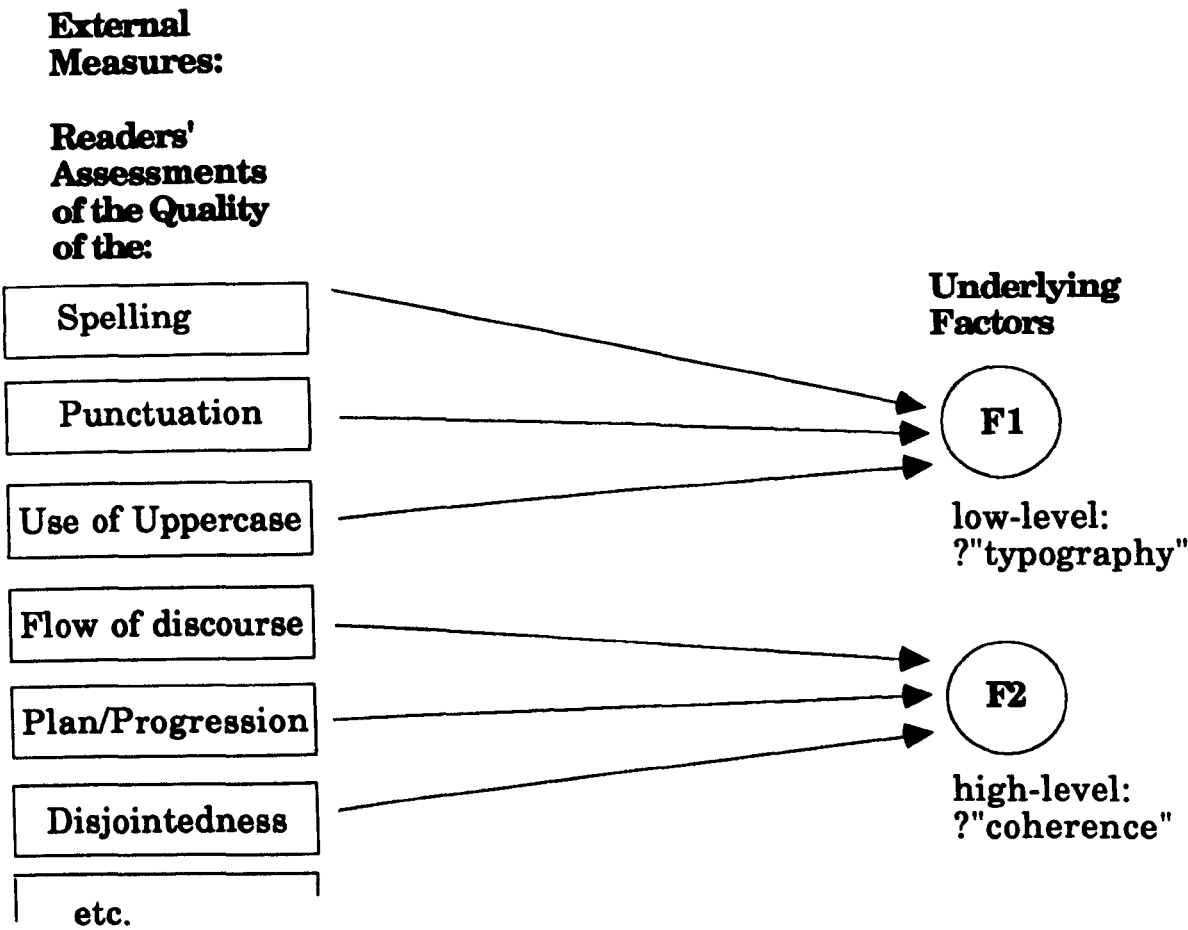


In this example, the variables on the left hand side of the diagram represent the reader's responses to each of the attitude questions. Similar types of variable are clustered to result in a smaller number of factors. Two things are of note here: firstly, the resulting factors are not correlated, a fact which makes them appropriate input to Discriminant Function Analysis (see next section); secondly, the new factors must be given mnemonic names by the analyst - these will then be the underlying dimensions, or determinants of the perceived success of a text. The way these determinants relate to its overall perceived quality can be investigated using certain regression techniques, discussed in the next section.

Factor Analysis could be used to benefit discourse theory in two ways: by clustering together similar kinds of external measure; and by clustering together similar kinds of internal measure.

For example, a Factor Analysis of external measures could be diagrammed

as follows:



In this way, Factor Analysis can show which sets of judges' impressions are similar and which are different. The above diagram shows how the following comments - 'didn't flow at all', 'disjointed' and 'no clear plan/progression evident' (see Appendix 8.4.1) - may all be addressing similar facets of the underlying concept of coherence. The reader is referred to Biber 1988 and Churchill 1991 for further technical details.

To sum up, then, Factor Analysis provides:

- a way of reducing complex data to more manageable size, specifically:
- a means of assessing the degree to which the linguistic variables (internal measures) inter-relate;
- a means of assessing the degree to which the informants' impressions (external measures) inter-relate.

15.4.2 Discriminant Function Analysis

In what follows a powerful statistical procedure is presented which may help disentangle the many different linguistic features of a text which collectively determine its perceived success.

Discriminant Function Analysis is an advanced multivariate statistical technique which can be used in a variety of different ways. Given a number of cases assumed to belong to a small number of non-overlapping groups (sets of judges' scores for the perceived success of an abstract, for example), the technique can indicate which of a certain set of continuous variables (sets of linguistic features, for example) are the best predictors of group membership. It is also possible to order and quantify those variables' contribution to the discrimination. As well as *describing* the determinants of group membership, Discriminant Function Analysis can also *predict*, on the basis of its model of known groupings and the distribution of the continuous variables, into which category a new, unknown case would most probably be classified. This means that, with a trustworthy model of perceived success trained upon many previously taken internal and external measures, it should be possible to predict, on the basis of its linguistic organisation, how successful a new text would be perceived to be.

Unfortunately, Discriminant Function Analysis is a particularly complex statistical procedure, involving some complex mathematics. However, the technique can be performed by computer: statistical packages that offer Discriminant Function Analysis include (SAS Institute Inc., 1982: 381 - 396), SYSTAT (SYSTAT Inc., 1989: 292 - 299), and SPSS (SPSS Inc., 1990: 1 - 42).

The technique is used in advertising, in the assessment of creditworthiness, financial risk analysis, and in marketing, where questions may include: which characteristics best differentiate heavy smokers from light?; why should some people prefer Persil over Fairy Liquid?; which particular features of a product best determine customer satisfaction?, and so on. For the discourse theorist, the technique can help to answer the question, which linguistic features of a text are the best predictors of how that text is valued by readers.

Although lexical texture, grammatical intricacy and Theme have been concentrated upon in the analysis phase, it is not the contention of this work that success can wholly be accounted for by reference to these features alone. However, Discriminant Function Analysis can be used to provide a numerical

assessment of their contribution.

The technique could be used in a number of different ways. Judges' responses could be assessed *individually* to produce a number of quite possibly different models of success, each representing one person's point of view. Alternatively, a *more general* view of perceived success could be produced by using a text's overall rank, as has been done in Part Four. It would be particularly interesting to investigate how and why judges differ in their perceptions of success.

Further, Discriminant Function Analysis offers a way not only of comparing and contrasting judges' responses along single variables (for example, 'Ms Smith likes lots of marked themes, whereas Ms Jones prefers texts with unmarked themes'), but also enables comparison to be made between variables ('Mr Brown doesn't seem bothered either way about marked or unmarked themes; he's more interested in the way clauses are combined').

Two different kinds of variable are investigated in Discriminant Function Analysis: most usually, there is one single dependent, or grouping, variable and there are a number of independent, or predicting variables.

Assuming judges were asked to rank five abstracts according to how successful they believed them to be, the dependent variable 'success' would consist of five separate groups. In other words, there would be five different possible success scores for an abstract, ranging from 1 (the abstract perceived to be the most successful) to 5 (the abstract perceived to be the least successful).

The independent variables, on the other hand, could consist of sets of judges' quantitative evaluations of the different drivers of success collected by means of the type of questionnaire described above. Once again, these would range from 1 ('Much better than I expected') to 5 ('Much worse than I expected'). The relationship between the two types of variable can be represented diagrammatically as follows:

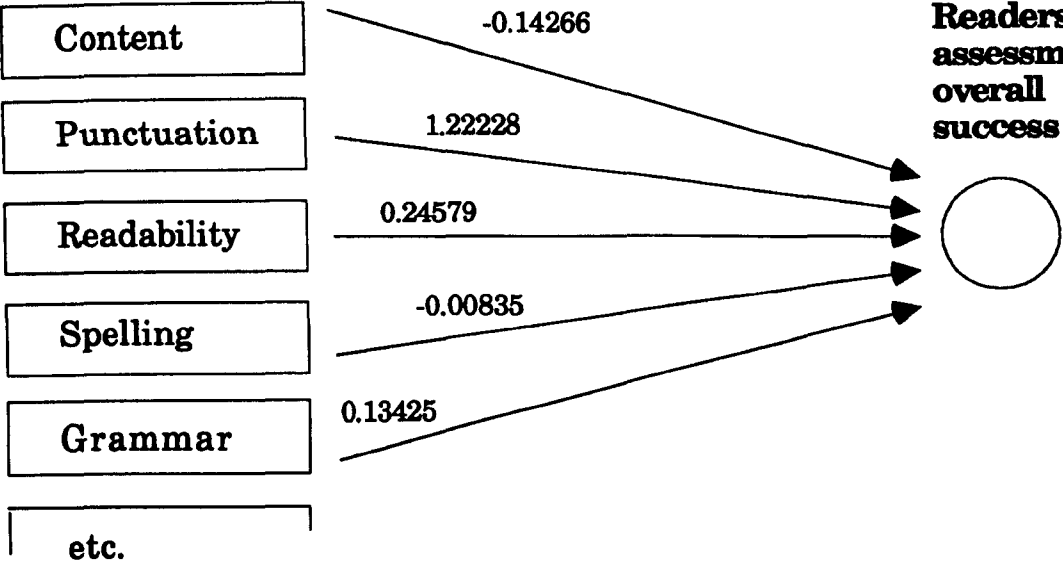
Independent Variables.

Readers' assessment of the quality of the:

- Content
- Punctuation
- Readability
- Spelling
- Grammar
- etc.

Dependent Variable.

Readers' assessment of overall success



Without going into any mathematical detail, Discriminant Function Analysis can provide one 'standardised canonical discriminant function coefficient' for each of the independent variables. These coefficients can be thought of as the weights which attach to each of the arrows in the above diagram, and give an indication of the relative predictive power of each variable. In the diagram above, for example, the (invented) coefficients suggest that the perceived success of the text is mainly determined by the quality of the punctuation. This illustrates how Discriminant Function Analysis could be used to reconcile external measures with other external measures.

However, it could also be used to reconcile external measures with internal measures:

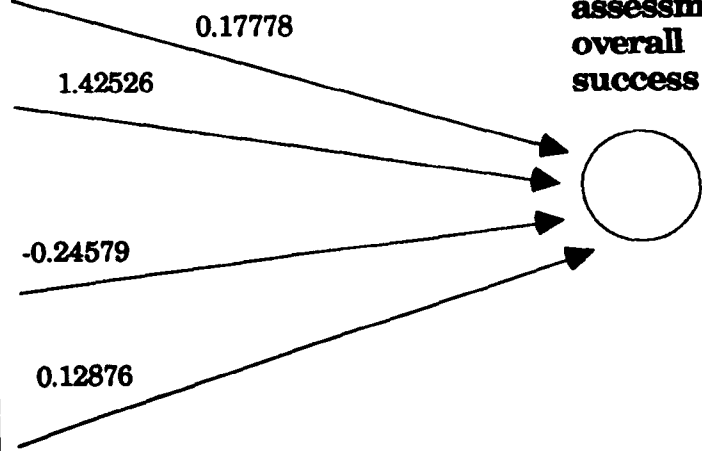
Independent Variables.

Linguistic measures:

- Average LD
- Grammatical Intricacy
- Proportion of Informational to Non-Informational Themes
- Amount of Enhancement
- etc.

Dependent Variable.

Readers' assessment of overall success



In this diagram, the (invented) coefficients suggest that the perceived success of the text is mainly determined by how grammatically intricate it is perceived to be. Different dependent variables could also be used. The impact of the different linguistic features on readers' assessments of readability of the text could be assessed, for example, showing which particular linguistic phenomena most affect the ease with which the text is read.

Experimental scenarios such as these show how Research Question 5 (Chapter 14 above) could be answered in further investigations. Unfortunately, there were too few cases in the data collected for the present study (see Chapters 6 and 7 above) to enable Discriminant Function Analysis; as a rule of thumb, at least 10 - 20 cases should be used for each independent variable, so when there are 4 predicting variables, there should be a minimum of between 40 - 80 cases. However, the type of data which was collected *does* constitute suitable input. Asking informants to *rank* a small number of abstracts ensures that there is a satisfactory spread in the data (i.e. there is something to discriminate between), and that abstracts can be assigned to groups relatively safely with comparatively little information being lost. Where different sets have different numbers of abstracts, these should be standardised: the suggestion here is to either assign the highest ranking abstract to group 1, the lowest to group 3, and those in the middle to group 2, or

to assign the middle abstract (or where the set contains an even number of abstracts, the middle two) to group 2, every abstract ranked higher to group 1, and every abstract ranked lower to group 3. Either way Discriminant Function Analysis will generate more reliable results if the number of groups is kept small. Further it is suggested that, although the number of cases should be large, informants should be presented with no more than six or seven texts to rank at a time.

In conclusion, the researcher in discourse theory will need access to three different kinds of expertise.

Firstly, external measures must be collected in as natural and grounded a way as possible. These should consist of a large sample of subjective opinions from the readers of the texts the researcher is interested in. It is suggested that the best kind of person to carry out this activity is someone with a background in marketing research, since marketing researchers should be best qualified to collect the relevant people's attitudes.

Secondly, internal measures must be collected. These should be chosen and applied bearing in mind the nature and scope of the external measures. Therefore, collection of the external measures must precede collection of the internal measures. The internal measures will derive from those types of linguistic analysis which promise to explain the external measures. It is suggested that the best kind of person to carry out this activity is someone with a background in textlinguistics, since textlinguists should be best qualified to suggest possible linguistic frameworks from which useful internal measures can be derived.

Thirdly, the two types of measure must be reconciled. This should be done by rigorously testing explicitly stated hypotheses, and by comparing and contrasting the predictive power of the various linguistic variables. Given a set of linguistic measures taken from a new text, ideally one would like to be able to predict how successful that text would be perceived to be. This is possible; it is suggested that the best kind of person to carry out this activity is someone with a background in multivariate statistics.

The greatest advances in discourse theory will be made by carefully combining these three types of skill.

References

- Akamatsu, C.T. (1988) 'Summarizing stories: the role of instruction in text structure in learning to write' American Annals of the Deaf 133, 4 pp 294 - 302
- ANSI (American National Standards Institute) (1977) 'American national standard for writing abstracts' IEEE Transactions on Professional Communication Vol. PC-20, No. 4 pp 252 - 254
- Arena, L.A. (1982) 'The language of corporate attorneys' in Di Pietro, R.J. (ed.) Linguistics and the Professions: Proceedings of the Second Annual Delaware Symposium on Language Studies Norwood, N.J.: Ablex
- Armbruster, B.B., Anderson, T.H. and Ostertag, J. (1987) 'Does text structure/summarization instruction facilitate learning from expository text?' Reading Research Quarterly 22, 3 pp 331 - 346
- ASTIA (Armed Services Technical Information Agency) (1959) Guidelines for Cataloguing and Abstracting Department of Defense, USA
- Backman, J. (1978) 'Reading comprehension and perceived comprehensibility of lexical density at discourse and sentence level' Scandinavian Journal of Educational Research 22, 1 pp 1 - 9
- Baxendale, P.B. (1958) 'Machine-made index for technical literature - an experiment' IBM Journal of Research and Development 2 pp 354-361
- Becker, H.S. (1970) Sociological Work: Method and Substance Chicago: Aldine
- Berry, M. (1975) Introduction to Systemic Linguistics, Volume 1: Structures and Systems London: Batsford
 - Berry, M. (1982) 'Review of Halliday 1978' Nottingham Linguistic Circular 11 pp 64 - 94
 - Berry, M. (1989) 'Thematic options and success in writing' in Butler, C.S., Cardwell, R.A. and Channell, J. (eds.) Language & Literature - Theory & Practice: a Tribute to Walter Grauberg Nottingham: Nottingham Linguistic Circular Special Issue
 - Berry, M. (1989b) 'They're all out of step except our Johnny: a discussion of motivation (or the lack of it) in Systemic Linguistics' Occasional Papers in Systemic Linguistics 3 pp 5 - 67

• Berry, M. (1990) 'Thematic analysis and stylistic preferences' Paper read to the 17th International Systemic Congress, University of Stirling, 3rd - 7th July, 1990

✓ • Berry, M. (forthcoming) 'Towards a study of generic choices and discourse demands' To appear in Discourse in Society: Systemic-Functional Perspectives (Volume 3 of Meaning and Choice in Language: Studies for Michael Halliday) To be published by Ablex.

• Berry, R. (1986) How to Write a Research Paper Oxford: Pergamon Press

Biber, D. (1986) 'Spoken and written textual dimensions in English: resolving the contradictory findings' Language 62, 2 pp 384 - 414

Biber, D. (1988) Variation Across Speech and Writing Cambridge: Cambridge University Press

Borko, H. and Bernier, C.L. (1975) Abstracting Concepts and Methods New York: Academic Press

Borko, H. and Chatman, S. (1963) 'Criteria for acceptable abstracts: a survey of abstracters' instructions' American Documentation April 1963 pp 149 - 160

Botel, M., Dawkins, J. and Granowsky, A. (1973) 'A syntactic complexity formula' in MacGinitie, W.H. (ed.) Assessment Problems in Reading Newark, Del.: International Reading Association

Bradac, J.J., Desmond, R.J. and Murdock, J.I. (1977) 'Diversity and density: lexically determined evaluative and informational consequences of linguistic complexity' Communication Monographs 44 pp 273 - 283

Brewer, J. and Hunter, A. (1989) Multimethod Research: a Synthesis of Styles London: Sage

Brown, A.L. and Day, J.D. (1983) 'Macrorules for summarizing texts: the development of expertise' Journal of Verbal Learning and Verbal Behavior 22 pp 1 - 14

Brown, A.L., Day, J.D. and Jones, R.S. (1983) 'The development of plans for summarizing texts' Child Development 54 pp 968 - 979

Brown, G. and Yule, G. (1983) Discourse Analysis Cambridge: Cambridge University Press

Bulmer, M. (1984a) 'Introduction: problems, theories and methods in sociology - (how) do they interrelate?' in Bulmer (ed.) Sociological Research Methods: an Introduction, 2nd edition. London: Macmillan

Bulmer, M. (1984b) 'Concepts in the analysis of qualitative data' in Bulmer (ed.) Sociological Research Methods: an Introduction, 2nd edition. London: Macmillan

- Butler, C. (1985a) Systemic Linguistics: Theory and Applications London: Batsford Academic
- Butler, C. (1985b) Computers in Linguistics Oxford: Blackwell
- Butt, D. (1984) 'Perceiving as making in the poetry of Wallace Stevens' Nottingham Linguistic Circular 13 pp 124 - 145
- Campbell, L.J. and Holland, V.M. (1982) 'Understanding the language of public documents because readability formulas don't' in Di Pietro, R.J. (ed.) Linguistics and the Professions: Proceedings of the Second Annual Delaware Symposium on Language Studies Norwood, N.J.: Ablex
- Carrell, P.L. (1982) 'Cohesion is not coherence' TESOL Quarterly 16, 4 pp 479 - 488
- Carter, R. (ed. 1982) Linguistics and the Teacher London: Routledge and Kegan Paul
- Carter, R. and Nash, W. (1983) 'Language and literariness' Prose Studies 6, 2 pp 123 - 141
- Chomsky, N. (1965) Aspects of the Theory of Syntax Cambridge, Mass.: The M.I.T. Press
- Christensen, F. (1968) 'The problem of defining a mature style' English Journal 57 pp 572 - 579
- Churchill, G.A., Jr. (1991) Marketing Research: Methodological Foundations Orlando, FL: Dryden Press
- Cicourel, A.C. (1964) Method and Measurement in Sociology New York: The Free Press
- Cleveland, D.B. and Cleveland, A.D. (1983) Introduction to Indexing and Abstracting Littleton, Colorado: Libraries Unlimited, Inc.
- Coleman, E.B. (1971) 'Developing a technology of written instruction: some determiners of complexity in prose' in Rothkopf, E.Z. and Johnson, P.E. (eds.) Verbal Learning Research and the Technology of Written Instruction New York: Teachers' College Press
- Collins (1990) Collins Cobuild English Grammar London: Collins ELT
- Collison, R.L. (1971) Abstracts and Abstracting Services Oxford: ABC Clio Press
- Cook, T.D. and Campbell, D.T. (1979) Quasi-Experimentation: Design and Analysis Issues for Field Settings Chicago: Rand McNally

- Couture, B. (1986) 'Effective ideation in written text: a functional approach to clarity and exigence' in Couture, B. (ed.) Functional Approaches to Writing: Research Perspectives Norwood, N.J.: Ablex
- Cremmins, E.T. (1982) The Art of Abstracting Philadelphia: ISI Press
- Crismore A. (1983) 'Metadiscourse: what it is and how it is used in school and non-school social science texts' (Tech. Rep. No.273) Urbana-Champaign: Center for the Study of Reading, University of Illinois
- Crismore, A. (1984) 'The rhetoric of text books: metadiscourse' Journal of Curriculum Studies 16 pp 279 - 296
- Cullingford, R.E. (1978) 'Script application: computer understanding of newspaper stories' Research Report #116, Yale University Department of Computer Science, New Haven, Connecticut.
- Danes, F. (1974) 'Functional sentence perspective and the organisation of text' in Danes, F. (ed.) Papers on Functional Sentence Perspective Prague: Academica
- Davies, F. (1988) 'Reading between the lines: thematic choice as a device for presenting writer viewpoint in academic discourse' The ESpecialist 1 - 2 pp 173 - 200. N.B. page references in the text refer to a pre-publication draft
- DeJong, G. (1982) 'An overview of the FRUMP system' in Lehnert, W.G. and Ringle, M.H. (eds.) Strategies for Natural Language Processing Hillsdale, N.J.: Erlbaum
- Denzin, N.K. (1970) The Research Act: a Theoretical Introduction to Sociological Methods New York: McGraw Hill
- Dixon, E. (1972) 'Syntactic indices and student writing performance' Elementary English 49 pp 714 - 716
- Dixon, J. (1967) Growth through English Reading: National Association for the Teaching of English
- Downing, A. (1990) 'On topical theme in English' Paper read to the 17th International Systemic Congress, University of Stirling, 3rd - 7th July, 1990
- Downing, A. (forthcoming) 'The discourse function of initial elements in written English: a modification of the Hallidayan approach' Mimeo; to be published in Word as 'An alternative approach to Theme: a systemic-functional approach'
- Drury, H. (1989a) 'The language of summaries: a systemic linguistic description: a comparative analysis of ESL and native speaker summaries of a journal article' Unpublished M.A. dissertation, Department of Linguistics, University of Sydney.

Drury, H. (1989b) 'The use of systemic linguistics to describe student summaries at university level' Paper presented to the 16th International Systemic Congress, June 1989

Dyer, M.G. (1983) In-depth Understanding: A Computer Model of Integrated Processing and Memory for Narrative Comprehension Cambridge, Mass.: MIT Press

Edmundson, H.P. and Wyllys, R.E. (1961) 'Automatic abstracting and indexing- survey and recommendations' Communications of the ACM 4,5 pp 226 - 234

• Enkvist, N.E. (1973) "'Theme dynamics" and style: an experiment' Studia Anglica Posnaniensa 5 pp 127 - 135

• Enkvist, N.E. (1978) 'Some aspects of applications of text linguistics' in Kohonen, V. and Enkvist, N.E. (eds.) Text Linguistics, Cognitive Learning and Language Teaching Turku, Finland: University of Turku

ERIC (Educational Resources Information Center) (1980) ERIC Processing Manual Rules and Guidelines for the Acquisition, Selection, and Technical Processing of Documents and Journal Articles by the Various Components of the ERIC Network Bethesda: ERIC Processing and Referencing Facility

Fichtner, E.G. (1980) 'Measuring syntactic complexity: the quantification of one factor in linguistic difficulty' Die Unterrichtspraxis 13, 1 pp 67 - 75

Fielding, N.G. and Fielding, J.L. (1986) Linking Data: the Articulation of Qualitative and Quantitative Methods in Social Research Sage University Paper Series on Qualitative Research Methods, Volume 4. Beverly Hills, CA: Sage

• Firbas, J. (1964) 'On defining the theme in Functional Sentence Analysis' TLP 1 pp 267 - 280

• Firbas, J. (1987) 'On the delimitation of the theme in functional sentence perspective' in Dirven, R. and Fried, V. (eds.) Functionalism in Linguistics Amsterdam: Benjamins

Fodor, J.A. and Garrett, M. (1967) 'Some syntactic determinants of sentential complexity' Perception and Psychophysics 2 pp 289 - 296

Fox, B.A. (1987) Discourse Structure and Anaphora Cambridge: Cambridge University Press

• Francis, G. (1989) 'Thematic selection and distribution in written discourse' Word 40, 1 - 2 pp 201 - 221

• Francis, G. (1990) 'Theme in the daily press' Occasional Papers in Systemic Linguistics 4 pp 51 - 87

Fries, P.H. (1983) 'On the status of theme in English: arguments from discourse' in Petöfi, J.S. and Sözer, E. (eds.) Micro and Macro Connexity of Texts Hamburg: Helmut Buske Verlag

Fries, P.H. (1986) 'Language features, textual coherence and reading' Word 37, 1 - 2 pp 13 - 29

Fries, P.H. (1987) 'Review of the Deakin University Press Language Education Series. Frances Christie, general editor. Victoria: Deakin University Press' Word 38, 3 pp 216 - 221

• Fries, P.H. (forthcoming) 'Towards a discussion of the flow of information in written text' to appear in Gregord, M. (ed.) Relations and Functions within and around Language Amsterdam: John Benjamins

• Fries, P.H. and Fair, J. (1978) 'The structure of texts: a preliminary report' Studia Anglica Posnaniensia 10 pp 95 - 104

• Fries, P.H. and Francis, G. (1990) 'Exploring Theme: problems for research' Handout from a workshop held at the Seventeenth International Systemic Congress, Stirling, July 1990.

Fum, D., Guida, G. and Tasso, C. (1984) 'A propositional language for text representation' in Bara, B.G. and Guida, G. (eds.) Computational Models of Natural Language Processing Amsterdam: Elsevier Science Publishers B.V.

Fum, D., Guida, G. and Tasso, C. (1985a) 'A rule based approach to evaluating importance in descriptive texts' Proceedings of the Second Conference of the European Chapter of the Association for Computational Linguistics, March 1985, University of Geneva, Switzerland

Fum, D., Guida, G. and Tasso, C. (1985b) 'Evaluating importance: a step towards text summarization' Proceedings of the Ninth International Joint Conference on Artificial Intelligence, August 1985, Los Angeles, California

Gibson, T.R. (1988) 'Review of automatic text summarisation tools and techniques: spotters versus understanders' BT automatic text summarisation (Project 322) deliverable 5.

Gibson, T.R. (1989) 'Review of the Linguistics and Psychology literature' BT automatic text summarisation (Project 322) deliverable 12.

• Givón, T. (1979) On Understanding Grammar New York: Academic Press

Glaser, B. and Strauss, A. (1967) The Discovery of Grounded Theory Chicago: Aldine

Golub, L.S. and Kidder, C. (1974) 'Syntactic density and the computer' Elementary English 51 pp 1128 - 1131

Gorrell, R.M. (1983) 'How to Make Mulligan Stew: Process and Product Again' College Composition and Communication 34, pp 272 - 277

Graetz, N. (1985) 'Teaching EFL students to extract structural information from abstracts' in Ulijn, J.M. and Pugh, A.K. (eds.) Reading for Professional Purposes Leuven, Belgium: ACCO

Grammatik Mac (1990) Grammatik Mac User's Guide San Francisco, CA: Reference Software International

Granowsky, A. and Botel, M. (1974) 'Background for a new syntactic complexity formula' The Reading Teacher 28, 1 pp 31 - 35

Gray, D.E. (1970) So You Have to Write a Technical Report: Elements of Technical Report Writing Washington, D.C.: Information Resources Press

Greenbaum, T.L. (1988) The Practical Handbook and Guide to Focus Group Research Toronto: Lexington Books

Gregory, M. (1967) 'Aspects of varieties differentiation' Journal of Linguistics 3 pp 177 - 198

Grishman, R. (1986) Computational Linguistics: an Introduction Cambridge: Cambridge University Press

Hahn, U. (1984) 'Textual expertise in word experts: an approach to text parsing based on topic/comment monitoring' Universität Konstanz, Informationswissenschaft, TOPIC Project Report 10/84

Hahn, U. and Reimer, U. (1984) 'Computing text constituency: an algorithmic approach to the generation of text graphs' Universität Konstanz, Informationswissenschaft, TOPIC Project Report 8/84

• Halliday, M.A.K. (1967) 'Notes on transitivity and theme in English, Part 2' Journal of Linguistics 3 pp 177 - 274

• Halliday, M.A.K. (1970) 'Language structure and language function' in Lyons, J. (ed.) New Horizons in Linguistics Harmondsworth, Middx.: Penguin

• Halliday, M.A.K. (1973) Explorations in the Functions of Language London: Edward Arnold

• Halliday, M.A.K. (1978) Language as Social Semiotic: The Social Interpretation of Language and Meaning London: Edward Arnold

Halliday, M.A.K. (1985) An Introduction to Functional Grammar London: Edward Arnold

• Halliday, M.A.K. (1987) 'Spoken and written modes of meaning' in Horowitz, R. and Samuels, S.J. (eds.) Comprehending Oral and Written Language London: Academic Press, Inc.

Halliday, M.A.K. (1988) 'On the language of physical science' in Ghadessy, M. (ed.) Registers of Written English: Situational Factors and Linguistic Features London: Pinter

Halliday, M.A.K. (1989) Spoken and Written Language Oxford: Oxford University Press

Halliday, M.A.K. and Hasan, R. (1976) Cohesion in English London: Longman

Hammersley, M. and Atkinson, P. (1989) Ethnography: Principles in Practice London: Routledge

Harber, J.R. 'Syntactic complexity: a necessary ingredient in predicting readability' Journal of Learning Disabilities 12, 8 pp 437 - 443

Harris, M.D. (1985) Introduction to Natural Language Processing Reston, Virginia, USA: Reston Publishing Co., Inc.

Hartnett, C.G. (1980) Cohesion as a Teachable Measure of Writing Competence Ph.D. dissertation, Indiana University of Pennsylvania

Hartnett, C.G. (1982) 'How cohesion expresses thinking' Journal of the Linguistic Association of the Southwest 5, 3 - 4 pp 210 - 215

Hartnett, C.G. (1986) 'Static and dynamic cohesion: signals of thinking in writing' in Couture, B. (ed.) Functional Approaches to Writing: Research Perspectives London: Pinter

- Hasan, R. (1978) 'Text in the systemic-functional model' in Dressler, W.U. (ed.) Current Trends in Textlinguistics New York: de Gruyter

Hasan, R. (1984) 'Coherence and cohesive harmony' in Flood, J. (ed.) Understanding Reading Comprehension: Cognition, Language, and the Structure of Prose Newark, Delaware: International Reading Association

- Hasan, R. (1985/1989a) 'The structure of a text'. Reprinted in Halliday, M.A.K. and Hasan, R. Language, Context and Text: Aspects of Language in a Social-Semiotic Perspective Oxford: Oxford University Press (page references are to the 1989 edition)

Hasan, R. (1985/1989b) 'The texture of a text'. First appeared in 1980; reprinted in Halliday, M.A.K. and Hasan, R. Language, Context and Text: Aspects of Language in a Social-Semiotic Perspective Oxford: Oxford University Press (page references are to the 1989 edition)

Herman, J. L., Morris, L. L. and Fitz-Gibbon, C. T. (1987) Evaluator's Handbook London: Sage

Hill, S.S., Soppelsa, B.F. and West, G.K. (1982) 'Teaching ESL students to read and write experimental-research papers' TESOL Quarterly 16, 3 pp 333 - 347

- Hoey, M. (1979) 'Signalling in Discourse' (Discourse Analysis Monograph, no. 6) English Language Research, University of Birmingham
- Hoey, M. (1983) On the Surface of Discourse London: George Allen and Unwin
- Hoey, M. Draft MS: 'Patterns of Lexis in Text' Unpublished MS, University of Birmingham
- Huddleston, R. (1978) 'On classifying anaphoric relations: a review of M.A.K. Halliday and R. Hasan's "Cohesion in English"' Lingua 45, 3 - 4 pp 333 - 354
- ✓ Huddleston, R. (1988) 'Constituency, multi-functionality and grammaticalization in Halliday's Functional Grammar' Journal of Linguistics 24 pp 137 - 174
- ✓ Hudson, R. (1986) 'Systemic Grammar' Linguistics 24 pp 791 - 815
- Hunt, K.W. (1965) Grammatical Structures Written at Three Grade Levels National Council of Teachers of English Research Report No. 3, Champaign, Illinois: National Council of Teachers of English
- Hunt, K.W. (1971) 'Teaching syntactic maturity' in Perren, G.E. and Trim, J.L.M. (eds.) Applications of Linguistics Cambridge: Cambridge University Press
- Hunt, K.W. (1977) 'Early blooming and late blooming syntactic structures' in Cooper, C.R. and Odell, L. (eds.) Evaluating Writing: Describing, Measuring, Judging Urbana, Illinois: National Council of Teachers of English
- ISO 214 (1976) Documentation - Abstracts for Publications and Documentation International Organization for Standardization, first edition 1976-03-01
- Johns, A.M. (1985) 'Summary protocols of "underprepared" and "adept" university students: replications and distortions of the original' Language Learning 35, 4 pp 495 - 517
- Jones, J. (1988) 'Grammatical metaphor and technicality in academic writing' MA long essay, Department of Linguistics, University of Sydney
- Kalivoda, T. (1980) 'Learning to listen - what can be done?' Forum 18, 4 pp 2 - 7
- Kent, A. and Lancour, H. (eds. 1968) Encyclopedia of Library and Information Science Vol 1 New York: Marcel Dekker Inc.
- King, J.R., Biggs, S. and Lipsky, S. (1984) 'Students' self-questioning and summarizing as reading study strategies' Journal of Reading Behaviour 16, 3 pp 205 - 218
- Kintsch, W. (1974) The Representation of Meaning in Memory Hillsdale, N.J.: Erlbaum

Kintsch, W. and van Dijk, T.A. (1978) 'Toward a model of text comprehension and production' Psychological Review 85 pp 363 - 394

Kirk, J. and Miller, M.L. (1986) Reliability and Validity in Qualitative Research Sage University Paper Series on Qualitative Research Methods, Volume 1. Beverly Hills, CA: Sage

Klare, G.E. (1963) The Measurement of Readability Ames, Iowa: The Iowa State University Press

Kolodner, J. (1980) 'Retrieval and organizational strategies in conceptual memory: a computer model' Technical Report #187, New Haven: Yale University Department of Computer Science

✓ Kroll, B. (1977) 'Combining ideas in written and spoken English: a look at subordination and coordination' in Keenan, E. and Bennett, R. (eds.) Discourse Across Time and Space Los Angeles: University of Southern California

Kucera, H. and Francis, W.N. (1970) Computational Analysis of Present-day American English Providence: Brown University Press

Kuhlen, R. (1983a) 'Some similarities and differences between intellectual and machine text understanding for the purpose of abstracting' Universität Konstanz, Informationswissenschaft, TOPIC Project Report 7/83

Kuhlen, R. (1983b) 'A knowledge-based text analysis system for the graphically supported production of cascaded text condensates' Universität Konstanz, Informationswissenschaft, TOPIC Project Report 9/83

Labov, W. (1972a) 'The transformation of experience in narrative syntax' in Language in the Inner City: Studies in the Black English Vernacular Philadelphia: University of Pennsylvania Press

• Labov, W. (1972b) Sociolinguistic Patterns Philadelphia: University of Pennsylvania Press

Lachenmeyer, C.W. (1971) The Language of Sociology London: Columbia University Press

Lane, J.C. (1961) 'Digesting for a multicompany management audience' Journal of Chemical Documentation 1,2 pp 59 - 63

Lehnert, W.G. (1980) 'Question answering in natural language processing' in Bolc, L. (ed.) Natural Language Question Answering Systems München Wien: Carl Hanser Verlag

Lincoln, Y.S. and Guba, E.G. (1985) Naturalistic Inquiry London: Sage

Linnarud, M. (1975) 'Lexical density and lexical variation - an analysis of the lexical texture of Swedish students' written work' Studia Anglica Posnaniensia 7 pp 45 - 52

Linnarud, M. (1977) 'Some aspects of style in the source and the target language' Papers and Studies in Contrastive Linguistics 7 pp 85 - 94

Luhn, H.P. (1958) 'The automatic creation of literature abstracts' IBM Journal of Research and Development 2:2 April 1958 pp 159-65

• Lyons, J. (1968) Introduction to Theoretical Linguistics Cambridge: Cambridge University Press

Maizell, R.E., Smith, J.F. and Singer, T.E.R. (1971) Abstracting Scientific and Technical Literature London: John Wiley and Sons

Malinowski, B. (1922) Argonauts of the Western Pacific London: Routledge and Kegan Paul

Mandler, J.M. and Johnson, N.S. (1977) 'Remembrance of things parsed: story structure and recall' Cognitive Psychology 9 pp 111 - 151

Mann, W. (1984) 'Discourse structures for text generation' Proceedings of COLING 84, pp 367 - 375; Tenth International Conference on Computational Linguistics; 2 - 6 July 1984, Stanford University, California.

Marshall, C. and Rossman, G.B. (1989) Designing Qualitative Research London: Sage

✓ • Martin, J.R. (1984) 'Language, register and genre' in Christie, F. et al. (eds.) Children Writing: Reader ECT418 Language Studies, Geelong, Vic., Deakin University Press

Martin, J.R. (1985) Factual Writing: Exploring and Challenging Social Reality Victoria, Australia: Deakin University Press

✓ • Martin, J.R. (1986a) 'Intervening in the process of writing development' in Painter, C. and Martin, J. R. (eds.) Writing to Mean: Teaching Genres across the Curriculum (Occasional Papers Number 9) Applied Linguistics Association of Australia

✓ • Martin, J.R. (1986b) 'Grammaticalising ecology: the politics of baby seals and kangaroos' in Threadgold, T., Grosz, E.A., Kress, G. and Halliday, M.A.K. (eds.) Language - Semiotics - Ideology Sydney: Sydney Association for Studies in Society and Culture

✓ • Martin, J.R. (1988) 'Hypotactic recursive systems in English: toward a functional interpretation' in Benson, J. D. and Greaves, W. S. (eds.) Systemic Functional Approaches to Discourse: Selected Papers from the 12th International Systemic Workshop Norwood, N.J.: Ablex

Martin, J.R. and Peters, P. (1985) 'On the analysis of exposition' in Hasan, R. (ed.) Discourse on Discourse Workshop reports from The Macquarie Workshop on Discourse Analysis. Applied Linguistics Association of Australia, Occasional Papers Number 7

Martin, J.R. and Rothery, J. (1986) 'What a functional approach to the writing task can show teachers about "good writing"' in Couture, B. (ed.) Functional Approaches to Writing: Research Perspectives London: Pinter

• Matthiessen, C. and Thompson, S.A. (1988) 'The structure of discourse and "subordination"' in Haiman, J. and Thompson, S. A. (eds.) Clause Combining in Grammar and Discourse Amsterdam: John Benjamins

McHale, B. (1978) 'Free indirect discourse: a survey of recent accounts' Poetics and Theory of Literature 3 pp 249 - 287

McKeown, K.R. (1982) 'The TEXT system for natural language generation' Proceedings of the 20th Annual Meeting of the Association for Computational Linguistics

Miles, M.B. and Huberman, A.M. (1984) Qualitative Data Analysis: a Sourcebook of New Methods London: Sage

Morgan, J.L. and Sellner, M.B. (1980) 'Discourse and linguistic theory' in Spiro, R.J., Bruce, B.C., and Brewer, W.F. (eds.) Theoretical Issues in Reading Comprehension Hillsdale, New Jersey: Lawrence Erlbaum Associates

Mosenthal, J.H. and Tierney, R.J. (1984) 'Cohesion: problems with talking about text' Reading Research Quarterly 19, 2 pp 240 - 244

Mosenthal, P.B. (1985) 'Defining the expository discourse continuum: towards a taxonomy of expository text types' Poetics 14 pp 387 - 414

Munch, J.M. and Swasy, J.L. (1988) 'Rhetorical question, summarization frequency, and argument strength effects on recall' Journal of Consumer Research 15, 1 pp 69 - 76

Murdock, J.I. and Desmond, R.J. (1975) 'An investigation of oral and written messages: production vs. comprehension' Paper presented at the Annual Meeting of the Speech Communication Association, Houston

• Nesbitt, C. and Plum, G. (1988) 'Probabilities in a systemic-functional grammar: the clause complex in English' in Fawcett, R.P. and Young, D.Y. (eds.) New Developments in Systemic Linguistics, Volume 2: Theory and Application London: Pinter

Neuner, J.L. (1985) 'Text studies and teaching: lessons from cohesion analysis' The English Record 36, 4 pp 8 - 12

- Noël, D. (1989) 'The study of coherence relations: what is wrong with it and how it can be improved upon' Interface 3, 2 pp 137 - 145
- Nwogu, K.N. (1990) Discourse Variation in Medical Texts: Schema, Theme and Cohesion in Professional and Journalistic Accounts Monographs in Systemic Linguistics. Nottingham: Department of English Studies, University of Nottingham
 - Nwogu, K. and Bloor, T. (1989) 'Thematic progression in professional and popular medical texts' Paper presented to the 16th International Systemic Congress, June 1989
- Odell, L. (1983) 'Written products and the writing process' in Hayes, J.N. and Roth, P.A., Ramsay, J.R. and Foulke, R.D. (eds.) The Writer's Mind: Writing as a Mode of Thinking Urbana, Illinois: National Council of Teachers of English
- Olson, G.M., Mark, R.L. and Duffy, S.A. (1981) 'Cognitive aspects of genre' Poetics 10 pp 283 - 315
 - Oppenheim, A.N. (1966) Questionnaire Design and Attitude Measurement London: Heinemann
- Paice, C.D. (1981) 'The automatic generation of literature abstracts: an approach based on the identification of self-indicating phrases' in Oddy, R.N., Robertson, S.E., van Rijsbergen, C.J. and Williams, P.W. (eds.) Information Retrieval Research London: Butterworths
- Parsons, G. (1990) Cohesion and Coherence: Scientific Texts, A Comparative Study Monographs in Systemic Linguistics, Volume 1. Nottingham: Department of English Studies, University of Nottingham
- Parsons, G. (forthcoming) Measuring Coherence in English Texts: an Investigation into the Relation between Cohesion and Coherence Ph.D. dissertation, University of Nottingham
- Patton, M.Q. (1987) How to Use Qualitative Methods in Evaluation London: Sage
- Patton, M.Q. (1990) Qualitative Evaluation and Research Methods, Second Edition London: Sage
- Pearson, P.D. (1974) 'The effects of grammatical complexity on children's comprehension, recall, and conception of certain semantic relations' Reading Research Quarterly 10, 2 pp 155 - 192
- Perera, K. (1982) 'The assessment of linguistic difficulty in reading material' in Carter, R. (ed.) Linguistics and the Teacher London: Routledge and Kegan Paul

- Perfetti, C.A. (1969) 'Lexical density and phrase structure depth variables in sentence recognition' Journal of Verbal Learning and Verbal Behavior 8 pp 719 - 724
- Peterson, C. and McCabe, A. (1983) Developmental Psycholinguistics New York: Plenum Press
- Peterson, L.N. and French, L. (1988) 'Summarization strategies of hearing-impaired and normally hearing college students' Journal of Speech and Hearing Research 31, 3 pp 327 - 337
- Phillips, B. (1985) Sociological Research Methods: an Introduction Homewood, Illinois: The Dorsey Press
- Pollock, J.J. and Zamora, A. (1975) 'Automatic abstracting research at Chemical Abstracts Service' Journal of Chemical Information and Computer Sciences Vol 15 No 4 pp 226 - 232
- Popper, K.R. (1959) The Logic of Scientific Discovery London: Hutchinson
- Popper, K.R. (1976) Unended Quest: an Intellectual Autobiography Glasgow: Fontana
- Potter, J. and Wetherell, M. (1987) Discourse and Social Psychology London: Sage
- Pritchard, R.J. (1980) A Study of the Cohesive Devices in the Good and Poor Composition of Eleventh Grades Ph.D. dissertation, University of Missouri
- Quirk, R., Greenbaum, S., Leech, G. and Svartvik, J. (1972) A Grammar of Contemporary English London: Longman
- Quirk, R., Greenbaum, S., Leech, G., and Svartvik, J. (1985) A Comprehensive Grammar of the English Language London: Longman
- Ravelli, L. (1985) 'Metaphor, mode and complexity: an exploration of covarying patterns' BA thesis: Department of Linguistics, University of Sydney
- Reimer, U. (1984) 'A system-controlled multi-type specialization hierarchy' Universität Konstanz, Informationswissenschaft, TOPIC Project Report 11/84
- Reimer, U. and Hahn, U. (1983) 'A formal approach to the semantics of the data frame model' in IJCAI 1983, Proceedings of the 8th International Joint Conference on Artificial Intelligence, Karlsruhe, August 8 - 12, 1983 pp 337 - 339
- Reis, R. and Leone, P.E. (1987) 'Teaching reading and study skills to mildly handicapped learners: previewing and text summarization' Pointer 31, 2 pp 41 - 43

- Rie, E.D. and Yeh, J.W. (1983) 'The effect of language on abstracting skills in learning disabled children' Journal of Clinical Child Psychology 12, 1 pp 40 - 45
- Riesbeck, C. (1982) 'Realistic language comprehension' in Lehnert, W.G. and Ringle, M.H. (eds.) Strategies for Natural Language Processing Hillsdale, N.J.: Erlbaum
- Rimmon-Kenan, S. (1983) Narrative Fiction: Contemporary Poetics London: Methuen
- Rinehart, S.D., Stahl, S.A. and Erickson, L.G. (1986) 'Some effects of summarization training on reading and studying' Reading Research Quarterly 21, 4 pp 422 - 438
- Rumelhart, D.E. (1975) 'Notes on a schema for stories' in Bobrow, D. and Collins, A. (eds.) Representation and Understanding New York: Academic Press
- Rumelhart, D.E. (1977) 'Understanding and summarizing brief stories' in La Berge, D. and Samuels, J. (eds.) Basic Processes in Reading Comprehension Hillsdale, N.J.: Erlbaum
- Sampson, G. (1980) Schools of Linguistics: Competition and Evolution London: Hutchinson
- * Sandulescu, C.-G. (1976) 'Theory and practice in analysing discourse' in Nickel, G. (ed.) Proceedings of the Fourth International Congress of Applied Linguistics, Volume 1 pp 349 - 365 Stuttgart: HochschulVerlag
- SAS Institute Inc. (1982) SAS™ User's Guide: Statistics Cary, N.C.: SAS Institute Inc.
- Schaeffer, K.W., Parsons, O.A. and Errico, A.L. (1988) 'Abstracting deficits and childhood conduct disorder as a function of familial alcoholism' Alcoholism: Clinical and Experimental Research 12, 5 pp 617 - 618
- Schank, R.C. (1972) 'Conceptual dependency: a theory of natural language understanding' Cognitive Psychology 3 pp 552-631
- Schank, R.C. and Abelson, R. (1977) Scripts, Plans, Goals and Understanding Hillsdale, N.J.: Erlbaum
- Schank, R.C. and Burstein, M. (1985) 'Artificial Intelligence: modelling memory for language understanding' in van Dijk, T.A. (ed.) Handbook of Discourse Analysis, Volume 1: Disciplines of Discourse New York: Academic Press
- Schank, R.C. and Riesbeck, C.K. (1981) Inside Computer Understanding: Five Programs plus Miniatures Hillsdale, N.J.: Erlbaum

- Schank, R.C., Goldman, N., Rieger, C. and Riesbeck, C. (1973) 'Margie: memory, analysis, response generation, and inference in English' Proceedings of the 3rd International Joint Conference on Artificial Intelligence, August, Stanford, pp 255-61
- Schank, R.C., Goldman, N., Rieger, C. and Riesbeck, C. (1975) 'Inference and paraphrase by computer' Journal of the ACM 22 pp 309 - 328
- Schluroff, M. (1982) 'Pupil responses to grammatical complexity of sentences' Brain and Language 17 pp 133 - 145
- Sherrard, C. (1986) 'Summary writing: a topographical study' Written Communication 3,3 pp 324 - 343
- Siegel, S. (1956) Nonparametric Statistics for the Behavioral Sciences London: McGraw-Hill
- Silverman, D. (1985) Qualitative Methodology and Sociology Aldershot, Hants.: Gower
- Sinclair, J. McH. and Coulthard, R.M. (1975) Towards an Analysis of Discourse: the English Used by Teachers and Pupils Oxford: Oxford University Press
- Sjoberg, G. and Nett, R. (1968) A Methodology for Social Research London: Harper and Row
- Speight, F.Y. (ed. 1967) Guide for Source Indexing and Abstracting of the Engineering Literature New York: Engineers Joint Council
- SPSS Inc. (1990) SPSS Advanced Statistics™ User's Guide Chicago, Ill.: SPSS Inc.
- Spurlin, J.E., Dansereau, D.F., O'Donnell, A. and Brooks, L.W. (1988) 'Text processing: effects of summarization frequency on text recall' Journal of Experimental Education 56, 4 pp 199 - 202
- Stainton, C. (forthcoming (a)) 'Genre and genre study: a review' To be published as 'What is this thing called Genre?' in Nottingham Working Papers
- Stainton, C. (forthcoming (b)) 'So why do I state "In this paper I will discuss ..."?' to be published in Nottingham Working Papers
- Stainton, C. (forthcoming (c)) 'Theme: comparisons and contrasts' to be published in Nottingham Working Papers
- Stanley, R.M. (1984) 'The recognition of macrostructure: a pilot study' Reading in a Foreign Language 2 pp 156 - 68

- Stein, N.L. and Glenn, C.G. (1979) 'An analysis of story comprehension in elementary school children' in Freedle, R.O. (ed.) New Directions in Discourse Processing II Norwood, N.J.: Ablex
- Strauss, A. and Corbin, J. (1990) Basics of Qualitative Research London: Sage
- Strauss, A.L. (1987) Qualitative Analysis for Social Scientists Cambridge: Cambridge University Press
- Stubbs, M. (1983) Discourse Analysis: the Sociolinguistic Analysis of Natural Language Oxford: Blackwell
- Stubbs, M. (1986) 'Lexical density: a technique and some findings' in Coulthard, M. (ed.) Talking about Text Discourse Analysis Monograph Number 13, University of Birmingham: English Language Research
- Swales, J.M. (1985) Episodes in ESP Oxford: Pergamon Press
- Swales, J.M. (1990) Genre Analysis: English in Academic and Research Settings Cambridge: Cambridge University Press
- SYSTAT Inc. (1989) SYSTAT™: The System for Statistics Evanston, Ill.: SYSTAT Inc.
- Tait, J.I. (1982) 'Automatic summarising of English texts' Technical Report No. 47, University of Cambridge Computer Laboratory, Cambridge
- Taylor, C.V. (1983) 'Structure and theme in printed school text' Text 3, 2 pp 197-228
- Templeton, J.F. (1987) Focus Groups: a Guide for Marketing and Advertising Professionals Chicago: Probus
- Thorndyke, P.W. (1977) 'Cognitive structures in comprehension and memory of narrative discourse' Cognitive Psychology 9 pp 77 - 110
- Turner, J. and Parsons, O.A. (1988) 'Verbal and nonverbal abstracting: problem-solving abilities and familial alcoholism in female alcoholics' Journal of Studies on Alcohol 49, 3 pp 281- 287
- Ure, J. (1971) 'Lexical Density and Register Differentiation' in Perren, G.E. and Trim, J.L.M. (eds.) Applications of Linguistics Cambridge: Cambridge University Press
- van Dijk, T.A. and Kintsch, W. (1983) Strategies of Discourse Comprehension New York: Academic Press
- Vande Kopple, W.J. (1985) 'Some exploratory discourse on metadiscourse' College Composition and Communication 36 pp 82 - 93

- Ventola, E. (1984) 'Orientation to social semiotics in foreign language teaching' Applied Linguistics 5 pp 275 - 286
- Vogel, S.A. (1985) 'Syntactic complexity in written expression of LD college writers' Annals of Dyslexia 35 pp 135 - 157
- Walker, L.A. (1985) 'An investigation of the effects on reading comprehension of teaching ninth graders to become independent summarizers of social studies materials' Dissertation, University of Columbia: Dissertation Abstracts International, Oct 1985 Vol 46 (4-A) pp 938 - 939
- Weil, B.H. (1970) 'Standards for writing abstracts' Journal of the American Society for Information Science 21, 5 pp 351 - 357
- Weil, B.H., Zarembek, I. and Owen, H. (1963) 'Technical Abstracting Fundamentals. Part II. Writing Principles and Practices' Journal of Chemical Documentation 3 pp 125 - 132
- Wilensky, R. (1978) 'Understanding goal-based stories' Research Report #140, Yale University Department of Computer Science, New Haven, Connecticut.
- Witte, S.P. and Cherry, R.D. (1986) 'Writing processes and written products in composition research' in Cooper, C.R. and Greenbaum, S. (eds.) Studying Writing: Linguistic Approaches London: Sage
- Witte, S.P. and Faigley, L. (1981) 'coherence, cohesion, and writing quality' College Composition and Communication 32, 2 pp 189 - 204
- Wolcott, H.F. (1990) Writing Up Qualitative Research Sage University Paper Series on Qualitative Research Methods, Volume 20. Beverly Hills, CA: Sage
- Yngve, V. (1960) 'A model and a hypothesis for language structure' in Proceedings of the American Philosophical Society, Volume 104, Number 5
- Youngman, M.B. (1982) Designing and Analysing Questionnaires Rediguide 12, Guides in Educational Research, Nottingham: Nottingham University School of Education
- Zeithaml, V.A., Parasuraman, A. and Berry, L.L. (1990) Delivering Quality Service: Balancing Customer Perceptions and Expectations New York: The Free Press
- Zora, S. and Johns-Lewis, C. (1989) 'Lexical density in interview and conversation' York Papers in Linguistics 14 pp 89 - 100

